REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE 20MAY02	3. REPORT TYPE AND DATES COVERED Non-peer reviewed reprint

4. TITLE AND SUBTITLE
Multitarget Moments and Their Application to Multitarget Tracking

5. FUNDING NUMBERS
DAAG55-98-C-0039

6. AUTHOR(S)
Ronald P.S. Mahler

2002 MAY 22 AM 9:00

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lockheed Martin Tactical Systems, 3333 Pilot Knob Road, Eagan MN 55121

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

37629-EL
.12

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The concept of a "statistical moment" has played a fundamental role in practical single-target tracking. The optimal tracking approach, the recursive Bayes filter, propagates the entire posterior density through time. Because this filter is computationally daunting, most practical single-target tracking approaches assume that signal-to-noise ratio is large enough that the posterior is approximately characterized by its low-order moments. For example, the alpha-beta-gamma filter propagates the first-order moment (the posterior expectation) whereas the extended Kalman filter (EKF) additionally propagates a second-order moment (the posterior covariance). Until recently, the possibility of an analogous multitarget approach seems to have been ignored---apparently for lack of a systematic statistical foundation for multitarget problems. In two recent papers, I introduced multitarget moment statistics of arbitrary order and developed a Bayes filtering theory for the first-order multitarget moment, the "probability hypothesis density (PHD)." In this paper I continue this line of investigation. I will describe a preliminary implementation of the first-order filter. I will introduce the concept of multitarget posterior covariances of arbitrary order. Using them, I will show how a suitable extension of the finite-set statistics (FISST) multisensor-multitarget differential calculus can be used to construct multitarget statistical analogs of the EKF.

14. SUBJECT TERMS
data fusion, random sets, point processes, multitarget tracking, statistical moments, spatial Poisson processes

15. NUMBER OF PAGES
34

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

20030605 053

Multitarget Moments and Their Application to Multitarget Tracking

Ronald Mahler
Lockheed Martin Tactical Systems, MS U2S25, 3333 Pilot Knob Rd, Eagan MN 55121, USA

ABSTRACT

The concept of a "statistical moment" has played a fundamental role in practical single-target tracking. The optimal tracking approach, the recursive Bayes filter, propagates the entire posterior density through time. Because this filter is computationally daunting, most practical single-target tracking approaches assume that signal-to-noise ratio is large enough that the posterior is approximately characterized by its low-order moments. For example, the alpha-beta-gamma filter propagates the first-order moment (the posterior expectation) whereas the extended Kalman filter (EKF) additionally propagates a second-order moment (the posterior covariance). Until recently, the possibility of an analogous multitarget approach seems to have been ignored—apparently for lack of a systematic statistical foundation for multitarget problems. In two recent papers, I introduced multitarget moment statistics of arbitrary order and developed a Bayes filtering theory for the first-order multitarget moment, the "probability hypothesis density (PHD)." In this paper I continue this line of investigation. I will describe a preliminary implementation of the first-order filter. I will introduce the concept of multitarget posterior covariances of arbitrary order. Using them, I will show how a suitable extension of the finite-set statistics (FISST) multisensor-multitarget differential calculus can be used to construct multitarget statistical analogs of the EKF.

1. INTRODUCTION

Progress in single-target tracking has relied upon the fundamental concept of the *moment statistics* of a random track-vector. The optimal approach to single-target tracking, the recursive Bayes filter, propagates the entire posterior density through time. Because real-time implementation of the Bayes filter is extremely challenging, most practical single-target tracking approaches assume that SNR is large enough that the posterior is approximately characterized by its low-order moments. For example, the alpha-beta-gamma filter propagates the first-order moment (the posterior expectation) whereas the extended Kalman filter (EKF) additionally propagates a second-order moment (the posterior covariance).

Until recently, the possibility of an analogous approach has apparently been overlooked as a multitarget tracking strategy. The reason for this has been, apparently, the lack of a systematic engineering statistics for multitarget problems. Even more surprisingly, this has been the case *even though a rigorous statistical foundation for multi-object problems, called* point process theory, *has been in existence for decades*. Specifically: How would one even go about defining first- and second-order multitarget moment statistics, let alone developing a rigorous Bayes filtering theory for them? As we shall see in a moment, naive approaches to defining such statistics fail. Consequently, one must turn to more systematic, theoretically-grounded approaches.

In two recent papers [27], [17], I resumed a line of investigation initiated in the book *Mathematics of Data Fusion* [9, pp. 168-170]. I introduced multitarget moment statistics of arbitrary order called "factorial-moment densities" and, in particular, the first-order moment also known as the "Probability Hypothesis Density" or "PHD". I developed a Bayes filter for the PHD, using a random-set formulation of point process theory called "finite-set statistics (FISST)." I showed that this first-moment filter is a multitarget statistical analog of the alpha-beta-gamma filter. That is, it is based on the assumption that signal-to-noise ratio is so high that the multitarget posterior density

Author e-mail and telephone: ronald.p.mahler@lmco.com; 651-456-4819.

is approximately characterized by its PHD. (Equivalently, it is based on the assumption that the multitarget track-process is approximately Poisson.) In this paper I continue this line of investigation. I will describe a preliminary implementation of the PHD filter. I will introduce the concept of *multitarget posterior covariances* of arbitrary order, and show how the FISST multi-target differential and integral calculus can be used to construct these moments. I will also show how a suitable extension of this calculus can potentially be used to construct multi-target statistical analogs of the Kalman filter (KF) and extended Kalman filter (EKF).

All of this, in turn, depends on a firm grounding in point process theory. Progress in single-sensor, single-object tracking has also been greatly facilitated by the existence of a systematic, rigorous, and yet practical *engineering statistics* that supports the development of new concepts in this field. By "engineering statistics," I mean the vast body of applied mathematical techniques surrounding the following "Statistics 101" concepts that most tracking engineers learn as undergraduates: random vectors; probability-mass and probability-density functions; differential and integral calculus; statistical moments (expected value, covariance, etc.); optimal state estimators; optimal signal-processing filters; and so on. Given the importance of such concepts in the single-sensor, single-object realm, one would expect that multisensor, multi-object tracking would already rest upon a similarly systematic, rigorous, and yet practical engineering statistics. This has not been the case despite the existence of point process theory, and there appear to be two major reasons for the gap. First, theoretical development in the multitarget tracking community has been focused primarily on immediate engineering applications rather than on systematic, over-arching foundations. Second, neither of the primary mathematical formulations of point process theory—*random measure theory* [6], [15] and *stochastic geometry* [48], [30]—have been well-suited for reduction to the practical "Statistics 101" form favored by tracking engineers. This is partly because random measure theory and stochastic geometry both tend to be somewhat impenetrable to even mathematically sophisticated engineers.* It is also partly due to the fact that point process theory has been directed not at tracking but, rather, towards those applications which gave birth to it: cueing theory, image signal processing, statistical theories of gases, liquids, and particles, etc. What has been missing has been an "engineering friendly" formulation of point process theory: "finite-set statistics"(FISST), the multisensor-multitarget engineering statistics that I introduced in 1994 . FISST is essentially a judicious, engineering-oriented distillation of point process and related concepts drawn from stochastic geometry, random measure theory, modern statistical physics, and expert-systems theory (see section 2). FISST is "engineering friendly" in that it is *geometric* (it treats multi-object systems as visualizable *random images*) and in that it *preserves and extends the "Statistics 101" formalism that tracking engineers already understand.*

1.1. Statistical Moments in Single-Target Tracking

With the clarity of hindsight, one can recognize that single-target tracking has taken the following conceptual (if not actual historical) trajectory. If real-time computational tractability were not an issue, optimal single-target trackers would be implementations of the following *Bayesian discrete-time recursive nonlinear filtering equations* [13], [42], [2, pp. 373-377], [14, p. 174]:

$$f_{k+1|k}(\mathbf{x}_{k+1}|Z^k) = \int f_{k+1|k}(\mathbf{x}_{k+1}|\mathbf{x}_k)\, f_{k|k}(\mathbf{x}_k|Z^k)d\mathbf{x}_k \tag{1}$$

$$f_{k+1|k+1}(\mathbf{x}_{k+1}|Z^{k+1}) = \frac{f_k(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})\, f_{k+1|k}(\mathbf{x}_{k+1}|Z^k)}{f_{k+1}(\mathbf{z}_{k+1}|Z^k)} \tag{2}$$

$$\hat{\mathbf{x}}_{k|k} = \arg\sup_{\mathbf{x}}\, f_{k|k}(\mathbf{x}|Z^k) \tag{3}$$

where

(1) $\mathbf{x}_k$ is the target state variable at time-step k and $\mathbf{z}_k$ is the observed measurement at time-step k;

*For example, at the 1998 GTRI/ONR Workshop on Tracking and Sensor Fusion [41], A. Skorokhod gave a presentation on random measure-based multitarget tracking [38] that does not seem to have been well understood.

(2) $f_{k|k}(\mathbf{x}_k|Z^k)$ is the Bayes posterior distribution conditioned on the data-stream $Z^k = \{\mathbf{z}_1, ..., \mathbf{z}_k\}$;

(3) $f_k(\mathbf{z}|\mathbf{x})$ is the sensor likelihood function at time-step k;

(4) $f_{k+1|k}(\mathbf{x}_{k+1}|\mathbf{x}_k)$ is the target Markov transition density that models between-measurements target motion;

(5) $f_{k+1|k}(\mathbf{x}_{k+1}|Z^k)$ is the time-prediction of the posterior $f_{k|k}(\mathbf{x}_k|Z^k)$ to time-step $k+1$;

(6) $f_{k+1}(\mathbf{z}_{k+1}|Z^k) = \int f_k(\mathbf{z}_{k+1}|\mathbf{y}_{k+1})\ f_{k+1|k}\mathbf{y}_{k+1}|Z^k)d\mathbf{y}_{k+1}$ is the Bayes normalization constant; and

(7) $\hat{\mathbf{x}}_{k|k}$ is the Bayes-optimal maximum *a posteriori* (MAP) estimate of the target state at time-step k.

Real-time computational implementation of equations 1, 2, and 3 is extraordinarily difficult in general. So, instead, a historically important strategy has been to assume that signal-to-noise ratio is high enough that the higher-order moments of the evolving track-vector $\hat{\mathbf{x}}_{k|k}$ can be neglected. In this case, the first- and second-order moment (i.e., the posterior expectation vector and the posterior second-moment matrix)

$$\hat{\mathbf{x}}_{k|k} = \int \mathbf{x}\ f_{k|k}(\mathbf{x}|Z^k)d\mathbf{x}, \qquad \hat{M}_{k|k} = \int \mathbf{x}\mathbf{x}^T\ f_{k|k}(\mathbf{x}|Z^k)d\mathbf{x} \tag{4}$$

are always approximate sufficient statistics:

$$f_{k|k}(\mathbf{x}|Z^k) \cong f_{k|k}(\mathbf{x}|\hat{\mathbf{x}}_{k|k}, \hat{M}_{k|k}) = N_{\hat{P}_{k|k}}(\mathbf{x} - \hat{\mathbf{x}}_{k|k})$$

where "^{T}" denotes matrix transpose and where $N_{P_{k|k}}(\mathbf{x} - \hat{\mathbf{x}}_{k|k})$ denotes a multi-variate Gaussian distribution with covariance matrix $\hat{P}_{k|k} = \hat{M}_{k|k} - \hat{\mathbf{x}}_{k|k}\hat{\mathbf{x}}_{k|k}^T$. We can then propagate $\hat{\mathbf{x}}_{k|k}$ and $\hat{M}_{k|k}$ instead of the full distribution $f_{k|k}(\mathbf{x}|Z^k)$ using an extended Kalman filter. If SNR is so high that the second-order moment can be neglected as well, then

$$f_{k|k}(\mathbf{x}|Z^k) \cong f_{k|k}(\mathbf{x}|\hat{\mathbf{x}}_{k|k}) \tag{5}$$

and we can propagate $\hat{\mathbf{x}}_{k|k}$ alone using an even computationally simpler constant-gain Kalman filter—e.g., the α-β-γ filter.

1.2. The Failure of Naive Multitarget Moments

Given such hindsight, someone new to multitarget tracking might surmise that the field has, at the conceptual if not the historical level, followed a similar trajectory. That is, he or she might presume that one would (1) begin with a theoretically systematic framework, namely propagation of a *multitarget posterior density* $f_{k|k}(X|Z^{(k)})$ using a multitarget analog of equations 1 and 2; and (2) transcend computational difficulties by assuming that SNR is high enough that one can propagate first- and/or second-order moments of the evolving multitarget track-set $X_{k|k}$, instead of the full distribution $f_{k|k}(X|Z^{(k)})$. In reality, multitarget tracking has taken a quite different course. The concept that multitarget tracking should be theoretically grounded on a multitarget analog of equations 1 and 2 is apparently itself of very recent vintage (see section 3.6). The idea of constructing lower-order statistical moments for a multitarget system, and using them as the basis for multitarget tracking algorithms, seems to have been overlooked entirely. This, in turn, appears to be attributable to the lack of a systematic engineering statistics for dealing with multitarget problems, similar to the statistics that has long been available for single-target problems.

In particular, how would one even go about defining first- and second-order multitarget moment statistics? It is easy to see that a naive approach fails [22, p. 41], [24], [25]. The most obvious and direct way of defining the first-order moment of a random track-set $\Gamma_{k|k}$ at time-step k would be to compute its expected value $\mathrm{E}[\Gamma_{k|k}]$. Suppose that we are in one dimension and that $f_{k|k}(\emptyset)$ is the probability that $\Gamma_{k|k} = \emptyset$ (no targets), $f_{k|k}(x)$ is the likelihood of the event $\Gamma_{k|k} = \{x\}$ (single track with state x), $f_{k|k}(x_1, x_2)$ is the likelihood of the event $\Gamma_{k|k} = \{x_1, x_2\}$ (two targets with states x_1, x_2), and so on. If the expectation exists, it must have the form

$$\mathrm{E}[\Gamma_{k|k}] = \emptyset \cdot f_{k|k}(\emptyset) + \int \{x\} \cdot f_{k|k}(x)dx + \int \{x_1, x_2\} \cdot f_{k|k}(x_1, x_2)dx_1dx_2 + ...$$

Such an integral cannot even be defined unless, at minimum, the multitarget state space is a vector space—in particular, unless it has a concept of addition/subtraction. But how does one add the zero-target state $\emptyset$ to a single-target state $\{x\}$? Or a single-target state $\{x\}$ to a two-target state $\{x_1, x_2\}$? As a specific example, let:

$$f_{k|k}(X) = \begin{cases} 1/2 & \text{if} \quad X = \emptyset \\ \frac{1}{2}N_{\sigma^2}(x-1) & \text{if} \quad X = \{x\} \\ 0 & \text{if} \quad |X| \geq 2 \end{cases}$$

where $N_{\sigma^2}(x)$ is a normal distribution with a variance σ^2 that has units km^2. The expectation would then be

$$\mathrm{E}[\Gamma_{k|k}] = \emptyset \cdot f(\emptyset) + \int x\, f(x)dx = \frac{1}{2}(\emptyset + 1km)$$

Notice that, at minimum, we have a units-mismatch problem: we are asked to add the unitless quantity $\emptyset$ to the quantity 1 *km*. Even if we assume that the continuous variable x is discrete so that this problem disappears, we still must add the quantity $\emptyset$ to the quantity 1. If $\emptyset + 1 = \emptyset$ then $1 = 0$, which is impossible. If $\emptyset + 1 = 1$ then $\emptyset = 0$, so the same mathematical symbol represents two different states (the no-target state $\emptyset$ and the single-target state $x = 0$). The same problem occurs if we define $\emptyset + a = b_a$ for *any* real numbers a, b_a since then $\emptyset = b_a - a$.

These difficulties persist even if we assume that the number of targets is known *a priori*. For example, suppose that we have a two-target system with two-target distribution

$$f_{k|k}(x_1, x_2) = \frac{1}{2}N_{\sigma^2}(x_1 - \bar{x}_1)N_{\sigma^2}(x_2 - \bar{x}_2) + \frac{1}{2}N_{\sigma^2}(x_2 - \bar{x}_1)N_{\sigma^2}(x_1 - \bar{x}_2)$$

where the targets are highly separated: σ^2 is small and $\bar{x}_1$ is much larger than $\bar{x}_2$. Our intuition leads us to believe that the multitarget expectation of this distribution should be $\mathrm{E}[\Gamma_{k|k}] = \{\bar{x}_1, \bar{x}_2\}$. The naive expectation of the distribution $f_{k|k}(x_1, x_2)$, however, yields $\mathrm{E}[\Gamma_{k|k}] = \frac{1}{2}(\bar{x}_1 + \bar{x}_2)$. This is because it does not take account of the fact that $f_{k|k}(x_1, x_2)$ is the distribution of a *multitarget* system and, therefore, that its symmetry $f_{k|k}(x_1, x_2) = f_{k|k}(x_2, x_1)$ provides no information about this system other than the fact that it is multitarget.

1.3. First-Order Multitarget Moments and Filtering

By a *track-valued multitarget expectation* I mean a *set of specific tracks* of the form $\mathrm{E}[\Gamma] = \{\hat{\mathbf{x}}^1_{k|k}, ..., \hat{\mathbf{x}}^{\hat{n}}_{k|k}\}$. The proper definition of the track-valued multitarget expectation $\mathrm{E}[\Gamma_{k|k}]$ of a random track-set is a problem that currently resists solution (see section 3.5). As a result, one has no choice but to construct multitarget moments by first specifying some function ϕ that transforms multitarget state-sets $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ into elements $\phi(X)$ of some suitably well-behaved vector space. The transformation ϕ should be one-to-one and it should transform set-theoretic operations into corresponding vector-algebra operations—for example, $h(X \cup Y) = h(X) + h(Y)$ whenever $X \cap Y = \emptyset$. In this case we can compute *indirect* first-order moments of the form $\mathrm{E}[h(\Gamma_{k|k})]$ that will themselves be elements of this vector space. Two obvious candidates are the following [9, p. 179]:

$$h(\{\mathbf{x}_1, ..., \mathbf{x}_n\}) = \delta_{\mathbf{x}_1} + ... + \delta_{\mathbf{x}_n}, \qquad h(\{\mathbf{x}_1, ..., \mathbf{x}_n\}) = \Delta_{\mathbf{x}_1} + ... + \Delta_{\mathbf{x}_n}$$

where $\delta_{\mathbf{x}}(\mathbf{y})$ is the Dirac delta density concentrated at $\mathbf{x}$ and where $\Delta_{\mathbf{x}}$ is its corresponding probability-mass function, i.e. the Dirac measure $\Delta_{\mathbf{x}}(S) = \int_S \delta_{\mathbf{x}}(\mathbf{y})d\mathbf{y} = 1$ (if $\mathbf{x} \in S$) and $= 0$ (otherwise). In the first case $\mathrm{E}[h(\Gamma_{k|k})]$ will be a density function $\hat{D}_{k|k}(\mathbf{y})$ and in the second case it will be the measure $\hat{D}_{k|k}(S) = \int_S \hat{D}_{k|k}(\mathbf{y})d\mathbf{y}$ corresponding to this density.

As it turns out, $\hat{D}_{k|k}(\mathbf{y})$ and $\hat{D}_{k|k}(S)$ have long been recognized as useful statistical moments in point process theory, where they are variously known as the *expectation*, *mean*, or *first factorial-moment density resp. measure*. I first described them four years ago in the book *Mathematics of Data Fusion* [9, p. 168-170]. There, I also showed

that $\hat{D}_{k|k}(\mathbf{y})$ is the same thing as the *probability hypothesis density (PHD)*, a multitarget tracking and evidence-accumulation technique proposed in 1993 by M.C. Stein and C.L. Winter [46]. In two recent papers [27], [17] I derived a recursive Bayes filter for the PHD. This filter is based on a multitarget analogy with equation 5. That is, signal-to-noise ratio is assumed to be so high that the PHD is an approximate sufficient statistic,

$$f_{k|k}(X|Z^{(k)}) \cong f_{k|k}(X|\hat{D}_{k|k})$$

and consequently that we can propagate $\hat{D}_{k|k}$ rather than the full multitarget posterior distribution $f_{k|k}(X|Z^{(k)})$. This work is described more fully in section 3.

Engineers tend to react with puzzlement to the idea that the first moment of a random multitarget track-set $\Gamma_{k|k}$ is a density function, namely the PHD $\hat{D}_{k|k}(\mathbf{x}|Z^{(k)})$. What they naturally expect to see is a track-valued expectation. So, what exactly is the PHD? Intuitively speaking, just as the value of the probability density function $f_{\mathbf{X}}(\mathbf{x})$ of a continuous random vector $\mathbf{X}$ provides a means of describing the zero-probability event $\Pr(\mathbf{X} = \mathbf{x})$, so the PHD $\hat{D}_{k|k}(\mathbf{x}Z^{(k)})$ provides a means of describing the zero-probability event $\Pr(\mathbf{x} \in \Gamma_{k|k})$.† In other words, it is the *total posterior likelihood that the multitarget system contains a target that has state* $\mathbf{x}$, and so is the total probability density that accrues to the hypothesis $\mathbf{x}$ (and hence the name "probability hypothesis density"). Consequently, $\hat{D}_{k|k}(\mathbf{x}|Z^{(k)})$ will tend to have maxima approximately at the locations of the targets. Since the expected number of tracks $\hat{N}_{k|k} = \int \hat{D}_{k|k}(\mathbf{x}|Z^{(k)})d\mathbf{x}$ can be computed directly from the PHD, multitarget track-estimation can be accomplished by searching for the $[\hat{N}_{k|k}]$ highest peaks in the PHD, where $[x]$ denotes the nearest-integer function.

1.4. Multitarget Moments, Covariances, and Filtering

The PHD filter summarized in section 1.3 makes use only of the first-order multitarget moment statistic. It is natural to ask whether second-order statistics can be applied to multitarget systems. Can we propagate a second-order approximation of the multitarget posterior density (i.e., the PHD and a multitarget covariance) instead of the multitarget posterior itself? Can we expand the multitarget likelihood function and multitarget Markov density in some kind of "multitarget Taylor's series"? If so, we could potentially produce multitarget analogs of the Kalman filter and the EKF.

Another purpose of this paper is to continue the line of investigation begun with the PHD filter, and try to determine whether or not it is possible to construct multitarget analogs of the Kalman filter or extended Kalman filter (EKF). As a beginning, in 1997 I also identified multitarget moment statistics of higher order and showed that they are identical to well-known point process moments called "factorial moment densities" [9, p. 168-170].

Results in this direction, summarized in section 4, are preliminary. At this time, a computationally tractable "multitarget Kalman filter" based on second-order multitarget moment statistics does not appear to be imminent. However, it appears that it might be possible to develop a multitarget analog of the EKF, though the potential computational practicality of such a "multitarget EKF" remains to be seen. This work is reported in section 4.

1.5. Organization of the Paper

The paper begins, in section 2, with a short but systematic introduction to point processes (section 2.1), finite-set statistics (section 2.2), and the relationship between the two (section 2.3). Section 3 introduces the multitarget moment density function (section 3.1), the Bayes multitarget first-moment filter (sections 3.2 and 3.1), and a preliminary implementation of this "PHD filter" (section 3.4). The section concludes with a brief discussion of an open research question (how to define *track-valued* rather than *function-valued* multitarget first moments, section 3.5), and a summary of related research (section 3.6). Section 4 explores the possibility of extending this line of reasoning to second-order multitarget moments. Possible multitarget analogs of the Kalman filter and the EKF are

†One consequence of this fact is that the PHD is the same thing as a *fuzzy set of tracks* when target state space is discrete [9, p. 169], [27, p. 108], [17, Section 2.5].

explored in sections 4.2 and 4.4, respectively. Mathematical proofs have been relegated to section 5; conclusions may be found in section 6.

2. MULTI-OBJECT STATISTICS

The purpose of this section is to summarize the elements of point process theory needed for this paper. I describe point process theory in section 2.1. In sections 2.2 and 2.3 I summarize FISST and clarify its relationship with more traditional formulations of point process theory.

2.1. Point Processes and Functional Calculus

Randomly varying, multi-object ensembles are a basic feature of multitarget tracking. The ensemble Z of observations collected from a multitarget system is a set of randomly-varying observation vectors, the number of which is itself random. If we take a Bayesian viewpoint, then the ensemble X of target-tracks also is a set of randomly-varying objects (state vectors), the number of which is itself random. Point processes are the foundation for representing stochastic systems that, like these examples, consist of randomly varying objects that randomly vary in number. The purpose of this section is to provide a short summary of this theory. This summary is nonstandard, however, in that it borrows from and integrates the practice not only of statisticians and stochastic geometers, but also of physicists and of expert-systems theorists (most notably, I.R. Goodman).

In this section I define point processes (section 2.1.1) and two other basic concepts: the *probability generating functional* (section 2.1.2) and the *functional derivative* (section 2.1.3). Then I define the basic statistical descriptors of point processes: *Janossy densities* (the multi-object analogs of probability densities; section 2.1.4), *factorial moment densities* (the multi-object analogs of moments; section 2.1.5), and *factorial cumulant densities* (the multi-object analogs of covariances; section 2.1.6).

2.1.1. Simple point processes (finite random sets)

There are many ways of defining point processes. Let $\Delta_{\mathbf{y}}(S)$ be the "Dirac measure," defined as $\Delta_{\mathbf{y}}(S) = 1$ if $\mathbf{y} \in S$ and $\Delta_{\mathbf{y}}(S) = 0$ otherwise, where S is some space O of objects (e.g., states or measurements). Let $\delta_{\mathbf{y}}(\mathbf{x})$ denote the density function of $\Delta_{\mathbf{y}}(S)$, i.e. the Dirac delta function concentrated at the point $\mathbf{y}$. Finally, let Γ be a randomly varying finite set of objects and $\delta_\Gamma(\mathbf{x}) = \sum_{\mathbf{y}\in\Gamma} \delta_{\mathbf{y}}(\mathbf{x})$ the random density function which consists of the sum of the Dirac delta functions concentrated on these objects. Then

$$N_\Gamma(S) = \sum_{\mathbf{y}\in\Gamma} \Delta_{\mathbf{y}}(S) = \int_S \delta_\Gamma(\mathbf{x}) d\mathbf{x} = |\Gamma \cap S| \tag{6}$$

is called a "random counting measure" because it counts the randomly varying number $|\Gamma \cap S|$ of objects in Γ that are contained in the region $S \subseteq O$. All of the three following items have been known for decades [48, p. 100-102], [1], [39], [36] to be equivalent and interchangeable definitions of a (simple) *multi-dimensional point process*‡:

1. the random finite subset Γ;§
2. the random measure $N_\Gamma(S)$ or its corresponding random density function $\delta_\Gamma(\mathbf{x})$.

‡Point processes can also be defined as random variables on unions of vector spaces [6, p. 121]. Though such a formulation might seem intuitively appealing, it is so mathematically restrictive that it is rarely used.

§More general approaches allow the random subset Γ to be countably infinite but locally finite, i.e. $|\Gamma \cap K| < \infty$ for any bounded subset K. We will not follow this practice here.

Not all random multi-object systems of interest are point processes. Let $A_{\mathbf{y}} \in \mathbb{R}^+$ be a family of random positive real numbers indexed by $\mathbf{y} \in O$. Define the random density function

$$\delta_{A,\Gamma}(\mathbf{x}) = \sum_{\mathbf{y}\in\Gamma} A_{\mathbf{y}} \cdot \delta_{\mathbf{y}}(\mathbf{x}) \tag{7}$$

and its associated random measure $N_{A,\Gamma}(S) = \sum_{\mathbf{y}\in\Gamma} A_{\mathbf{y}} \Delta_{\mathbf{y}}(S)$. If the $A(\mathbf{x})$ were random positive integers then $N_{A,\Gamma}(S)$ would be a general point process; otherwise, it is called a *compound process.* Compound processes are nevertheless mathematically equivalent to simple point processes (finite random sets) on the product space $\mathbb{R}^+ \times O$ [15, pp. 5, 16, 22]. In particular, there is a sequence of notational correspondences

$$a_1\delta_{\mathbf{x}_1} + ... + a_n\delta_{\mathbf{x}_n} \Longleftrightarrow \delta_{(a_1,\mathbf{x}_1)} + ... + \delta_{(a_n,\mathbf{x}_n)} \Longleftrightarrow \{(a_1,\mathbf{x}_1), ..., (a_1,\mathbf{x}_1)\} \tag{8}$$

As we will see later (section 4.4), each of these three notations can be interpreted as a *finite collection of point object-clusters*, each such point object-cluster $(a, \mathbf{x})$ consisting of an average number a of objects all co-located at $\mathbf{x}$.

2.1.2. Probability generating functionals

In single-object statistics, the statistical behavior of a random number Y is often described by *generating functions* such as the characteristic function $\phi_Y(y)$ =E[e^{iyY}], the moment-generating function $M_Y(y)$ =E[e^{yY})], the factorial moment-generating function $G_Y(y)$ =E[y^Y)], and so on [4, p. 83], [6, Chapter 1]. These functions are called "generating functions" because probability functions and various kinds of moments can be generated from their iterated derivatives, e.g. $(d^jM_Y/dy^j)(0)$ =E[Y^j]. In like manner, the statistical behavior of the point process Γ, $N_\Gamma(S)$ or $\delta_\Gamma(\mathbf{x})$ is described by various kinds of generating *functionals*. My emphasis in this paper will be on a point-process analog of the factorial moment-generating function called the *probability generating functional.* Given a function h, this is (if it exists) the expected value of the product of all $h(\mathbf{x})$ taken over all elements $\mathbf{x}$ of Γ:

$$G_\Gamma[h] = \mathrm{E}\left[\prod_{\mathbf{x}\in\Gamma} h(\mathbf{x})\right] \tag{9}$$

In mathematical point process theory h must be a very restricted type of function, to ensure that the expectation exists for very general point processes Γ [6, p. 141], [48, p. 116], [36, p. 13]. However, I will follow the practice of physicists in assuming that Γ is sufficiently "nice" that $G_\Gamma[h]$ can be defined for more general functions h—in particular, functions that incorporate Dirac delta functions $\delta_{\mathbf{y}}(\mathbf{w})$ [40, p. 190]. The probability generating functional has the property that $G_{\Gamma_1,...,\Gamma_n}[h] = G_{\Gamma_1}[h] \cdots G_{\Gamma_n}[h]$ if $\Gamma_1, ..., \Gamma_n$ are statistically independent.

2.1.3. Functional derivatives

For any functional $F[h]$ and any function g define

$$\frac{\partial F}{\partial g}[h] = \lim_{\varepsilon\to 0} \frac{F[h+\varepsilon g] - F[h]}{\varepsilon}, \qquad \frac{\partial^n F}{\partial g_1 \cdots \partial g_n}[h] = \frac{\partial}{\partial g_1} \cdots \frac{\partial}{\partial g_n} F[h] \tag{10}$$

if the limit exists. In mathematics, this is called a "Gâteaux derivative." Like physicists, I assume that F is sufficiently nice that the transformation $g \longmapsto \frac{\partial F}{\partial g}[h]$ is continuous and linear for any fixed h. In this case it is also known as a "Frechét derivative" and can be computed using the usual "turn-the-crank" rules of freshman calculus. Provided that they exist, the iterated Frechét derivatives $(\partial^n F/\partial g_1 \cdots \partial g_n)[h]$ are linear in each argument $g_1, ..., g_n$. If Frechét derivatives of all orders exist, then F can be expanded in a Taylor's series [40, p. 190]

$$F[h] = \sum_{i=0}^{\infty} \frac{1}{i!} \frac{\partial^i F}{\partial (h-h_0)^i}[h_0] \qquad \text{where} \qquad \frac{\partial^i F}{\partial g^i}[h] = \frac{\partial^i F}{\underbrace{\partial g \cdots \partial g}_{\text{i times}}}[h] \tag{11}$$

Because of continuity and linearity and since $g = \int g \cdot \delta_{\mathbf{y}} d\mathbf{y}$ (that is, $g(\mathbf{w}) = \int g(\mathbf{y})\delta_{\mathbf{y}}(\mathbf{w})d\mathbf{y}$ for all $\mathbf{w}$), a physicist would write

$$\frac{\partial F}{\partial g}[h] = \int g(\mathbf{y}) \, \frac{\partial F}{\partial \delta_{\mathbf{y}}}[h] d\mathbf{y}$$

and then point out that general Frechét derivatives are completely specified by the Frechét derivatives

$$\overbrace{\frac{\delta^n F}{\delta \mathbf{y}_1 \cdots \delta \mathbf{y}_n}[h]}^{\text{FISST}} \quad = \quad \overbrace{\frac{\delta^n F}{\delta h(\mathbf{y}_1) \cdots \delta h(\mathbf{y}_n)}[h]}^{\text{physics}} \quad = \quad \overbrace{\frac{\partial^n F}{\partial \delta_{\mathbf{y}_1} \cdots \partial \delta_{\mathbf{y}_n}}[h]}^{\text{mathematics}} \tag{12}$$

These are known in the physics literature as *functional derivatives* [40, pp. 173-174]. Equation 12 displays three different notations for these derivatives: the ones preferred by mathematicians (rightmost) and physicists (center), and the less cumbersome abbreviated physics notation used in FISST and throughout this paper (leftmost).

2.1.4. Janossy densities

Given a list $\mathbf{y}_1, ..., \mathbf{y}_n$ of vectors, the iterated functional derivatives

$$j_{\Gamma,n}(\mathbf{y}_1, ..., \mathbf{y}_n) = \left. \frac{\delta^n G_\Gamma}{\delta \mathbf{y}_1 \cdots \delta \mathbf{y}_n}[h] \right|_{h=0} \tag{13}$$

are the *Janossy density functions* [6, pp. 122-123] of Γ. The family of Janossy densities of a point process Γ is the multi-object analog of the probability density $f_{\mathbf{Y}}(\mathbf{y})$ of a random vector $\mathbf{Y}$. That is, just as $f_{\mathbf{Y}}(\mathbf{y})$ describes the likelihood of occurrence of the zero-probability event $\mathbf{Y} = \mathbf{y}$, so $j_{\Gamma,n}(\mathbf{y}_1, ..., \mathbf{y}_n)$ describes the likelihood of occurrence of the zero-probability event $\Gamma = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$. The Janossy densities of simple point processes are completely symmetric in all arguments, vanish whenever $\mathbf{y}_i = \mathbf{y}_j$ for some $1 \leq i \neq j \leq n$ [6, pp. 134, Prop. 5.4.IV], and are jointly normalized in the sense that $\sum_{n=0}^{\infty} \frac{1}{n!} j_n(\mathbf{y}_1, ..., \mathbf{y}_n) = 1$. The relationship between the probability generating functional and the Janossy densities can be found by expanding $G_\Gamma[h]$ in a Taylor's series expansion (equation 11) around $h = 0$ [40, p. 190], [36, p. 15], [6, pp. 142, 148]:

$$G_\Gamma[h] = \sum_{n=0}^{\infty} \frac{1}{n!} \int h(\mathbf{y}_1) \cdots h(\mathbf{y}_n) \, j_{\Gamma,n}(\mathbf{y}_1, ..., \mathbf{y}_n) d\mathbf{y}_1 \cdots d\mathbf{y}_n \tag{14}$$

2.1.5. Factorial moment densities

In ordinary single-target statistics, moments of arbitrary order of a random vector $\mathbf{Y}$ with distribution $f_{\mathbf{Y}}(\mathbf{y})$ are the tensors (multi-argument linear functions):

$$m_{\mathbf{Y},n}(\mathbf{y}_1, ..., \mathbf{y}_n) = \int \langle \mathbf{y}_1, \mathbf{y} \rangle \cdots \langle \mathbf{y}_n, \mathbf{y} \rangle \, f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}$$

where $\langle \mathbf{y}_1, \mathbf{y}_2 \rangle$ is the usual scalar (i.e, dot) product of vectors. The multitarget moments of a point process Γ are multi-argument *nonlinear* functions. Given a list $\mathbf{y}_1, ..., \mathbf{y}_n$ of vectors, the iterated functional derivatives

$$m_{\Gamma,[n]}(\mathbf{y}_1, ..., \mathbf{y}_n) = \left. \frac{\delta^n G_\Gamma}{\delta \mathbf{y}_1 \cdots \delta \mathbf{y}_n}[h] \right|_{h=1} \tag{15}$$

are the *factorial moment densities* of the point process Γ [6, pp. 130, 222], [15, p. 11], [48, pp. 111, 116]. They can also be shown to be the expected value

$$m_{\Gamma,[n]}(\mathbf{y}_1, ..., \mathbf{y}_n) = \mathrm{E}\left[\sum_{\mathbf{w}_1 \neq ... \neq \mathbf{w}_n \in \Gamma} \delta_{\mathbf{y}_1}(\mathbf{w}_1) \cdots \delta_{\mathbf{y}_n}(\mathbf{w}_n) \right]$$

where the summation is taken over all n-tuples $(\mathbf{w}_1, ..., \mathbf{w}_n)$ of distinct objects [48, p. 38], [15, pp. 10-11].

Factorial moment measures are related to the Janossy densities and the probability generating functional by the equations [6, pp. 133, 142, 149, 222], [36, pp. 11, 17]:

$$m_{\Gamma,[n]}(\mathbf{y}_1, ..., \mathbf{y}_n) = \sum_{k=0}^{\infty} \frac{1}{k!} \int j_{\Gamma,[n+k]}(\mathbf{y}_1, ..., \mathbf{y}_n, \mathbf{w}_1, ..., \mathbf{w}_k) d\mathbf{w}_1 \cdots d\mathbf{w}_k \tag{16}$$

$$j_{\Gamma,[n]}(\mathbf{y}_1, ..., \mathbf{y}_n) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int m_{\Gamma,[n+k]}(\mathbf{y}_1, ..., \mathbf{y}_n, \mathbf{w}_1, ..., \mathbf{w}_k) d\mathbf{w}_1 \cdots d\mathbf{w}_k \tag{17}$$

$$G_\Gamma[1+h] = \sum_{k=0}^{\infty} \frac{1}{k!} \int h(\mathbf{y}_1) \cdots h(\mathbf{y}_k)\; m_{\Gamma,[k]}(\mathbf{y}_1, ..., \mathbf{y}_k) d\mathbf{y}_1 \cdots d\mathbf{y}_k \tag{18}$$

If $n = 0$ then $m_{\Gamma,[0]} = 1$; whereas if $n = 1$ then we get the PHD described in section 1.3: $m_{\Gamma,[1]}(\mathbf{y}) = D_\Gamma(\mathbf{y})$. It is clear from equation 16 that $m_{\Gamma,[n]}(\mathbf{y}_1, ..., \mathbf{y}_n) = 0$ if $\mathbf{y}_i = \mathbf{y}_j$ for some $i \neq j$ (since $j_{\Gamma,[n]}(\mathbf{y}_1, ..., \mathbf{y}_n)$ has the same property).

2.1.6. Factorial cumulant densities

In ordinary single-target statistics, the covariance moments of arbitrary order of a random vector $\mathbf{Y}$ are the tensors:

$$c_{\mathbf{Y},n}(\mathbf{y}_1, ..., \mathbf{y}_n) = \int \langle \mathbf{y}_1 - \hat{\mathbf{y}}, \mathbf{y} \rangle \cdots \langle \mathbf{y}_n - \hat{\mathbf{y}}, \mathbf{y} \rangle\; f_\mathbf{Y}(\mathbf{y}) d\mathbf{y}$$

where $\hat{\mathbf{y}}$ is the expected value of $\mathbf{Y}$. The multi-object analog of these quantities are the iterated functional derivatives

$$c_{\Gamma,[n]}(\mathbf{y}_1, ..., \mathbf{y}_n) = \left. \frac{\delta^n \log G_\Gamma}{\delta \mathbf{y}_1 \cdots \delta \mathbf{y}_n}[h] \right|_{h=1} \tag{19}$$

which are called the *factorial cumulant densities* of Γ [6, p. 146]. They are related to the probability generating functional by [6, pp. 224]

$$\log G_\Gamma[1+h] = \sum_{k=0}^{\infty} \frac{1}{k!} \int h(\mathbf{y}_1) \cdots h(\mathbf{y}_k)\; c_{\Gamma,[j]}(\mathbf{y}_1, ..., \mathbf{y}_k) d\mathbf{y}_1 \cdots d\mathbf{y}_k \tag{20}$$

If $n = 0$ then $c_{\Gamma,[0]} = 0$; and if $n = 1$ then $c_{\Gamma,[1]}(\mathbf{y}) = m_{\Gamma,[1]}(\mathbf{y}) = D_\Gamma(\mathbf{y})$. If $n = 2$ then we get the *factorial covariance density* [6, p. 146]:

$$\begin{aligned} c_\Gamma(\mathbf{y}_1, \mathbf{y}_2) &= c_{\Gamma,[2]}(\mathbf{y}_1, \mathbf{y}_2) = \left. \frac{\delta^2 \log G_\Gamma}{\delta \mathbf{y}_1 \delta \mathbf{y}_2}[h] \right|_{h=1} = -\frac{\delta G_\Gamma}{\delta \mathbf{y}_2}(1) \frac{\delta G_\Gamma}{\delta \mathbf{y}_1}(1) + \frac{\delta^2 G_\Gamma}{\delta \mathbf{y}_1 \delta \mathbf{y}_2}(1) \\ &= m_{\Gamma,[2]}(\mathbf{y}_1, \mathbf{y}_2) - D_\Gamma(\mathbf{y}_1) D_\Gamma(\mathbf{y}_2) \end{aligned} \tag{21}$$

Equation 21 can be generalized to statistics of all orders, since the factorial cumulant densities can be written as combinatorial sums of products of the factorial moment densities [6, p. 147].

2.2. Finite-Set Statistics (FISST)

As previously noted and as I explain more fully in this and the following sections, FISST is essentially a judicious, engineering-oriented distillation of point process and related concepts drawn from stochastic geometry, random measure theory, modern statistical physics, and expert-systems theory. The theoretical basis of FISST has been described in the book *Mathematics of Data Fusion* [9, Chapters 2, 4-8]. The engineering motivations underlying FISST have been summarized in the technical monograph *An Introduction to Multisource-Multitarget Statistics and*

Its Applications [22] and elsewhere [23], [21]. Rather than repeating this material here, I direct the reader to those sources. Instead, in this section I summarize the basic relationships between FISST and the (nonstandard) formulation of point process theory described in the previous section.

I begin with basic concepts: the belief-mass function (section 2.2.1), set derivative (section 2.2.2), multi-object density function (section 2.2.3), and set integral (section 2.2.4). The first three are the FISST counterparts of the probability generating functional, functional derivative, and Janossy densities, respectively. I then turn to engineering concepts: the multisensor-multitarget likelihood function (section 2.2.5), multitarget Markov transition densities (section 2.2.6), and multitarget posterior densities (section 2.2.7).

2.2.1. Belief-mass function

In single-sensor, single-target problems, tracking engineers typically do not use generating functions to describe a random vector $\mathbf{Y}$, but rather the probability-mass function $p_{\mathbf{Y}}(S) = \Pr(\mathbf{Y} \in S)$ or its density function $f_{\mathbf{Y}}(\mathbf{y})$. In like manner, FISST describes a point process Γ not with its probability generating functional (section 2.1.2) but rather its "belief-mass function" $\beta_\Gamma(S) = \Pr(\Gamma \subseteq S)$[¶]. The belief-mass function is the multi-object analog of the probability-mass function. Its value $\beta_\Gamma(S)$ is the total probability that all objects in Γ will be contained in S. The belief-mass function and probability generating functionals are closely related. For any closed subset S, let $\mathbf{1}_S(\mathbf{x})$ be the function defined by $\mathbf{1}_S(\mathbf{z}) = 1$ if $\mathbf{z} \in S$ and $\mathbf{1}_S(\mathbf{z}) = 0$ otherwise. Then the belief-mass function is the restriction of the probability generating functional of Γ to these characteristic functions (see section 5.1):

$$\beta_\Gamma(S) = G_\Gamma[\mathbf{1}_S]$$

2.2.2. Set derivative

Let $Y = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$. Then the FISST "set derivative" of a belief-mass function β_Γ of a finite-random set (point process) Γ is, if it exists,

$$\begin{aligned}
\frac{\delta \beta_\Gamma}{\delta \mathbf{y}}(S) &\triangleq \frac{\delta}{\delta \mathbf{y}}\beta_\Gamma(S) \triangleq \lim_{\lambda(E_{\mathbf{y}})\searrow 0} \frac{\beta_\Gamma(S \cup E_{\mathbf{y}}) - \beta_\Gamma(S)}{\lambda(E_{\mathbf{y}})} \\
\frac{\delta \beta_\Gamma}{\delta Y}(S) &\triangleq \frac{\delta^n \beta_\Gamma}{\delta \mathbf{y}_1 \cdots \delta \mathbf{y}_n}(S) \triangleq \frac{\delta}{\delta \mathbf{y}_1} \cdots \frac{\delta}{\delta \mathbf{y}_n}\beta_\Gamma(S) \\
\frac{\delta \beta_\Gamma}{\delta \emptyset}(S) &\triangleq \beta_\Gamma(S)
\end{aligned} \tag{22}$$

where $E_{\mathbf{y}}$ is a small region that converges to the singleton set $\{\mathbf{y}\}$ and $\lambda(E_{\mathbf{y}})$ is its hypervolume (i.e., Lebesgue measure). The set derivative can be thought of as a functional derivative. That is, if both the respective set derivatives and functional derivatives exist, then they are related by (see section 5.2):

$$\frac{\delta \beta_\Gamma}{\delta \mathbf{y}}(S) = \frac{\delta G_\Gamma}{\delta \mathbf{y}}[\mathbf{1}_S], \qquad \frac{\delta \beta_\Gamma}{\delta Y}(S) = \frac{\delta^n G_\Gamma}{\delta \mathbf{y}_1 \cdots \delta \mathbf{y}_n}[\mathbf{1}_S] = \frac{\partial^n G_\Gamma}{\partial \delta_{\mathbf{y}_1} \cdots \partial \delta_{\mathbf{y}_1}}[\mathbf{1}_S]$$

2.2.3. Multi-object density functions

Let Γ be a simple point process (i.e., random finite set). We construct the *multi-object density function* $f_\Gamma(X)$ of Γ by bundling together the Janossy densities of Γ (section 2.1.4) into a single density function defined on the space of finite random sets:

$$f_\Gamma(Y) = j_{\Gamma,n}(\mathbf{y}_1, ..., \mathbf{y}_n) = \frac{\delta^n G_\Gamma}{\delta \mathbf{y}_1 \cdots \delta \mathbf{y}_n}[0] = \frac{\delta^n \beta_\Gamma}{\delta \mathbf{y}_1 \cdots \delta \mathbf{y}_n}(\emptyset) = \frac{\delta \beta_\Gamma}{\delta Y}(\emptyset)$$

[¶]In stochastic geometry and theoretical statistics, $\beta_\Gamma(S)$ is known as a specific kind of "capacity measure." If Γ is nonempty (i.e., $\Pr(\Gamma \neq \emptyset) = 1$) then $\beta_\Gamma(S)$ is a "belief function" in the sense of the Dempster-Shafer theory. Because of this, the FISST formulation of point process theory encompasses Dempster-Shafer, fuzzy logic, and rule-based expert systems approaches.

where $Y = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$. That is, the multi-object density can be constructed as a set derivative. Also, $Y = \emptyset$ indicates that no object is present, $Y = \{\mathbf{y}\}$ indicates that one object with state $\mathbf{y}$ is present, $Y = \{\mathbf{y}_1, \mathbf{y}_2\}$ indicates that two objects with states $\mathbf{y}_1 \neq \mathbf{y}_2$ are present, and so on. The multi-object density $f_\Gamma(Y)$ has the form

$$\begin{aligned} f_\Gamma(\emptyset) &= \text{likelihood that no objects are present} \\ f_\Gamma(\{\mathbf{y}\}) &= \text{likelihood of one object } \mathbf{x} \\ &\vdots \\ f_\Gamma(\{\mathbf{y}_1, ..., \mathbf{y}_n\}) &= \text{likelihood of } n \text{ (distinct) objects } \mathbf{x}_1, ..., \mathbf{x}_n \end{aligned}$$

2.2.4. Set integrals

Given any function $f(Y)$ of a finite set variable Y, the *set integral* is defined by

$$\int_S f(Y)\delta Y = \sum_{j=0}^{\infty} \frac{1}{j!} \int_{\underbrace{S \times ... \times S}_{j \text{ times}}} f(\{\mathbf{y}_1, ..., \mathbf{y}_j\}) d\mathbf{y}_1 \cdots d\mathbf{y}_j \tag{23}$$

for any measurable set S, where by convention the $j = 0$ term in the infinite sum is $f(\emptyset)$. If $f = f_\Gamma$ for some point process Γ, then for each j the j^{th} term in this sum is the probability $p_{\Gamma,j} = \Pr(|\Gamma| = j)$ that Γ contains Γ objects. Given this, equation 14 can be rewritten as

$$G_\Gamma[h] = \int h^Y \cdot f_\Gamma(Y)\delta Y$$

where $h^Y = \prod_{\mathbf{y} \in Y} h(\mathbf{y})$. Consequently, if $f_\Gamma(Y)$ is the multi-object density function of Γ, then $\int f_\Gamma(Y)\delta Y = G_\Gamma[1] = 1$. Substituting $h = \mathbf{1}_S$ yields

$$\beta_\Gamma(S) = G_\Gamma[\mathbf{1}_S] = \int \mathbf{1}_S^Y f_\Gamma(Y)\delta Y = \int_S f_\Gamma(Y)\delta Y$$

The set integral is inverse to the set derivative, in the sense that

$$\beta(S) = \int_S \frac{\delta\beta}{\delta Y}(\emptyset)\delta Y, \qquad f(Y) = \left[\frac{\delta}{\delta Y}\int_S f(W)\delta W\right]_{S=\emptyset}$$

2.2.5. Multitarget likelihood functions

Let Σ_k be the random observation-set generated by a multitarget system with multitarget state X at time-step k and $\beta_k(S|X) = \Pr(\Sigma_k \subseteq S|X)$ its belief-mass function. Let $Z = \{\mathbf{z}_1, ..., \mathbf{z}_m\}$ be a particular such observation-set. Then the multitarget likelihood function $f_k(Z|X)$ of Σ_k is the multi-object density function of Σ_k. That is, it is constructed by bundling the family of Janossy densities of Σ_k into a single density function defined on finite sets Z of measurements. In particular, it is an iterated functional derivative of the probability generating functional of Σ_k [22, p. 30]:

$$f(Z|X) = j_{\Sigma_k,n}(\mathbf{z}_1, ..., \mathbf{z}_m) = \frac{\delta\beta_k}{\delta Z}(\emptyset|X)$$

2.2.6. Multitarget Markov transition densities

Likewise, let Let $\Gamma_{k+1|k}$ be the random track-set at time-step $k+1$ generated by a multitarget system with multitarget state X at time-step k. Let $\beta_{k+1|k}(S|X) = \Pr(\Gamma_{k+1|k} \subseteq S|X)$ be its belief-mass function and let $Y = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$ be a particular track-set at time-step $k+1$. Then the multitarget Markov density $f_{k+1|k}(Y|X)$ for $\Gamma_{k+1|k}$ is the multi-object density function of $\Gamma_{k+1|k}$—i.e., the family of Janossy densities of the point process $\Gamma_{k+1|k}$, bundled together into a single density defined on finite sets. In particular, it is a set derivative of the belief-mass function $\beta_{k+1|k}$ [22, p. 30]:

$$f_{k+1|k}(Y|X) = j_{\Gamma_{k+1|k},n}(\mathbf{y}_1, ..., \mathbf{y}_n) = \frac{\delta \beta_{k+1|k}}{\delta Y}(\emptyset|X)$$

2.2.7. Multitarget posterior densities

Let $Z^{(k)} = \{Z_1, ..., Z_k\}$ be a time-series of multisource-multitarget observation-sets Z_j collected from the randomly-varying track-set $\Gamma_{k|k}$ of a multitarget system at time-step k. The Janossy densities of this point process (random finite set) are, when bundled together into a single density function defined on all finite sets of target states, the "multitarget posterior density function":

$$f_{k|k}(X|Z^{(k)}) = j_{\Gamma_{k|k},n}(\mathbf{x}_1, ..., \mathbf{x}_n) = \frac{\delta \beta_{k|k}}{\delta X}(\emptyset|Z^{(k)})$$

where $\beta_{k|k}(S|Z^{(k)}) = \Pr(\Gamma_{k|k} \subseteq S)$ is the belief-mass function of $\Gamma_{k|k}$. In other words, $X = \emptyset$ indicates that no target is present, $X = \{\mathbf{x}\}$ indicates that one target with state $\mathbf{x}$ is present, $X = \{\mathbf{x}_1, \mathbf{x}_2\}$ indicates that two targets with states $\mathbf{x}_1 \neq \mathbf{x}_2$ are present, and so on. So,

$$\begin{aligned}
f_{k|k}(\emptyset|Z^{(k)}) &= \text{posterior likelihood that no targets are present} \\
f_{k|k}(\{\mathbf{x}\}|Z^{(k)}) &= \text{posterior likelihood of one target with state } \mathbf{x} \\
&\vdots \\
f_{k|k}(\{\mathbf{x}_1, ..., \mathbf{x}_n\}|Z^{(k)}) &= \text{posterior likelihood of } n \text{ targets with states } \mathbf{x}_1, ..., \mathbf{x}_n
\end{aligned}$$

2.3. FISST vs. Conventional Point Process Theory

Given the existence of these relationships between FISST and the more usual versions of point process theory, what is the advantage of using FISST? When I published *Mathematics of Data Fusion* in 1997, I argued [9, pp. 71-72, 170] that FISST offered engineers certain advantages over a multi-object statistics based on random measures. First, it is *explicitly geometric* in that the random variates in question are actual sets of observations (visualizable random images) rather than abstract integer-valued measures. Second, because FISST provides a systematic foundation for both expert systems theory (fuzzy logic, Dempster-Shafer theory, rule-based inference) *and* multisensor, multitarget estimation and filtering, it permits a systematic and mathematically rigorous integration of these two quite different aspects of information fusion. Third, systematic adherence to a random set perspective results in a formulation of point process theory that is nearly identical to the "Statistics 101" formalism with which tracking engineers are already familiar. In subsequent publications [22], [23], [21] I spelled out what this means in greater detail:

- just as single-sensor, single-target data can be modeled using a measurement model $\mathbf{Z}_k = h_k(\mathbf{x}_k, \mathbf{W}_k)$, so multitarget multisensor data can be modeled using a *multisensor-multitarget measurement model*—a randomly varying finite set $\Sigma_k = T_k(X_k) \cup C_k(X_k)$ where T_k and C_k indicate target-generated and clutter-generated observations, respectively [22, pp. 17-20];

- just as the single-sensor, single-target likelihood function $f_k(\mathbf{z}|\mathbf{x}_k)$ can be derived from the probability mass function $p_k(S|\mathbf{x}_k) = \Pr(\mathbf{Z}_k \in S|\mathbf{x}_k)$ of the measurement model via differentiation, so the true multitarget likelihood function $f_k(Z|X_k)$ can be derived from the belief-mass function $\beta_k(S|X_k) = \Pr(\Sigma_k \subseteq S|X_k)$ of the multisensor-multitarget measurement model, using the set derivative;
- just as single-target motion can be modeled using a motion model $\mathbf{X}_{k+1} = \Phi_k(\mathbf{x}_k, \mathbf{V}_k)$, the motion of multitarget systems can be modeled using a *multitarget motion model*—a randomly varying finite set $\Gamma_{k+1} = \Phi_k(X_k, \mathbf{V}_k)$ that takes account of target appearance and disappearance [22, pp. 21-23];
- just as the Markov transition density $f_{k+1|k}(\mathbf{x}_{k+1}|\mathbf{x}_k)$ can be derived from the probability mass function $p_{k+1|k}(S|\mathbf{x}_k) = \Pr(\mathbf{X}_{k+1} \in S|\mathbf{x}_k)$ of the motion model via differentiation, so the true multitarget Markov transition density $f_{k+1|k}(X_{k+1}|X_k)$ can be derived from $\beta_{k+1|k}(S|X_k) = \Pr(\Gamma_{k+1|k} \subseteq S|X_k)$, by applying the set derivative to the belief-mass function of the multitarget motion model;
- just as simple "turn the crank" rules exist for freshman differential and integral calculus, so similar turn the crank rules (sum, product, chain, power) exist for the FISST differential and integral calculus [22, pp. 31-32].

Finally, because FISST systematizes these "Statistics 101" parallels between single-object and multi-objects statistics, it provides a general engineering methodology for attacking multisource-multitarget data fusion problems:

- **Almost-Parallel Worlds Principle (APWOP):** *Nearly any single-sensor, single-target concept or algorithm can, in principle, be directly translated into a corresponding multisensor, multitarget concept or algorithm*

Since 1994, a long-standing illustration of the APWOP has been the concept of *multitarget information measures of effectiveness*—for example, the following multitarget generalization of the Kullback-Leibler cross-entropy [9, pp. 205-209; 295-312], [18], [20], [52], [19]:

single-sensor, single-target	$\Longrightarrow$	multisensor-multitarget
$K(f,g) = \int f(\mathbf{x}) \log\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x}$	$\Longrightarrow$	$K(f,g) = \int f(X) \log\left(\frac{f(X)}{g(X)}\right) \delta X$

Using the APWOP, we simply replace conventional statistical concepts on the left with their FISST multisensor, multitarget counterparts on the right. Specifically, the ordinary densities f, g on the left are replaced by the multitarget densities f, g on the right; and the ordinary integral $\int \cdot d\mathbf{x}$ on the left is replaced by a multitarget set integral $\int \cdot \delta X$ on the right.

3. MULTITARGET FIRST-MOMENT FILTERING

The purpose of this section is to: introduce the concept of *multitarget moments of all orders* (section 3.1); describe a multitarget filtering approach based on propagation of the first-order moment (the "probability hypothesis density" or PHD, sections 3.2 and 3.3); and summarize performance results for a preliminary implementation of this "PHD filter" (section 3.4). As previously noted, the PHD filter described in these sections is a statistical multitarget analog of the α-β-γ filter. It sidesteps the computationally intractable problem of propagating the multitarget posterior density by, instead, propagating only the first-order moment of that density. This is possible only under the assumption that signal-to-noise ratio is relatively high. The PHD filter must be implemented as a computational nonlinear filter, using analogs of equations 1 and 2. It tracks targets without making any attempt to associate reports with tracks. For more details, the reader is referred to recent papers [27], [17], [29], [7]. I conclude the section with an open research problem (how to define track-valued rather than function- or measure-valued expected values, section 3.5)—and a short history of Bayes recursive multitarget nonlinear filtering (section 3.6).

3.1. The Multitarget Moment Density Function

Recall that in section 2.1, I used an iterated functional derivative of the probability generating functional $G_{\Gamma_{k|k}}[h]$ of a random track-set (point process) $\Gamma_{k|k}$ to define the Janossy densities $j_{\Gamma_{k|k},n}(\mathbf{x}_1, ..., \mathbf{x}_n)$. Then in section 2.2 I bundled these densities together into a single *multitarget posterior density function* $f_{k|k}(X|Z^{(k)})$ and showed that it can be computed as an iterated *set derivative* of the belief-mass function $\beta_{k|k}(S)$ of $\Gamma_{k|k}$. In section 2.1 I also introduced the concept of the factorial moment densities of a point process. Applying equation 15 to the random track-set $\Gamma_{k|k}$, we likewise bundle the factorial moment densities $m_{\Gamma_{k|k},[n]}(\mathbf{x}_1, ..., \mathbf{x}_n)$ into a single density function $m_{k|k}(X|Z^{(k)})$ defined on a finite-set variable $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ and call it the *multitarget moment density function* of the track-set:

$$m_{k|k}(\emptyset|Z^{(k)}) = 1, \qquad m_{k|k}(X|Z^{(k)}) = m_{\Gamma_{k|k},[n]}(\mathbf{x}_1, ..., \mathbf{x}_n)$$

From equation 15 it follows that

$$m_{k|k}(X|Z^{(k)}) = \frac{\delta \beta_{k|k}}{\delta X}(S|Z^{(k)})\bigg|_{S=\mathbb{S}} = \frac{\delta \beta_{k|k}}{\delta X}(\mathbb{S}|Z^{(k)}) \tag{24}$$

where $\mathbb{S}$ is the entire state space. Thus the multitarget moment density can, like the multitarget posterior density, be computed from $\beta_{k|k}$ using iterated set derivatives. Likewise, equations 16, 17, and 18 become

$$\begin{aligned} m_{k|k}(X|Z^{(k)}) &= \int_{Y \supseteq X} f_{k|k}(Y|Z^{(k)})\delta Y = \int f_{k|k}(X \cup W|Z^{(k)})\delta W & (25) \\ f_{k|k}(X|Z^{(k)}) &= \int_{Y \supseteq X} (-1)^{|W|}\, m_{k|k}(Y|Z^{(k)})\delta Y = \int (-1)^{|W|}\, m_{k|k}(X \cup W|Z^{(k)})\delta W & (26) \\ G_{k|k}[1+h] &= \int h^X \cdot m_{k|k}(X|Z^{(k)})\delta X & (27) \end{aligned}$$

where $h^X = \prod_{\mathbf{x} \in X} h(\mathbf{x})$ given $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$. Equations 25 and 26 tell us that the multitarget posterior and multitarget moment densities are interchangeably computable from each other. Equation 25 additionally tells us that $m_{k|k}(X|Z^{(k)})$ is the marginal-posterior likelihood that, regardless of how many targets there may be in the multitarget system, exactly n of them have states $\mathbf{x}_1, ..., \mathbf{x}_n$.

In particular, the first moment density (or "probability hypothesis density," PHD [46], [9, pp. 168-169]) has the form

$$D_{k|k}(\mathbf{x}|Z^{(k)}) = m_{k|k}(\mathbf{x}|Z^{(k)}) = \int_{X \ni \mathbf{x}} f_{k|k}(X|Z^{(k)})\delta X = \int f_{k|k}(\{\mathbf{x}\} \cup Y|Z^{(k)})\delta Y$$

For any measurable subset $S \subseteq \mathbb{S}$ of state-vectors, the expected number of targets in S is [27, p. 106], [17, Theorem 2], [9, p. 169]

$$N_{k|k} = \mathrm{E}[\,|\Gamma_{k|k} \cap S|\,] = \int_S D_{k|k}(\mathbf{x}|Z^{(k)})d\mathbf{x} \tag{28}$$

This property characterizes the PHD uniquely. That is, if $g_k(\mathbf{x})$ is any other density which gives the expected number of targets in S when integrated over S, then it is (no matter how imaginative the name or notation one might assign to it) nothing else but the PHD. For, since $\int_S g_k(\mathbf{x})d\mathbf{x} = \int_S D_{k|k}(\mathbf{x}|Z^{(k)})d\mathbf{x}$ for all measurable S then $g_{k|k} = D_{k|k}$ almost everywhere. A typical example of a PHD is pictured in Figure 2.

3.2. Basic Idea Behind the PHD Filter

The theoretically optimal foundation for multitarget detection, tracking, and identification is the following generalization of the recursive Bayes filter equations 1, 2, and 3:

$$f_{k+1|k}(X_{k+1}|Z^{(k)}) = \int f_{k+1|k}(X_{k+1}|X_k)\, f_{k|k}(X_k|Z^{(k)})\delta X_k \tag{29}$$

$$f_{k+1|k+1}(X_{k+1}|Z^{(k+1)}) = \frac{f_k(Z_{k+1}|X_{k+1})\, f_{k+1|k}(X_{k+1}|Z^{(k)})}{f_{k+1}(Z_{k+1}|Z^{(k)})} \tag{30}$$

$$\hat{X}^{JoM}_{k+1|k+1} = \arg\sup_X \frac{c^{|X|}}{|X|!} f_{k+1|k+1}(X|Z^{k+1}) \tag{31}$$

where [22, pp. 47-49], [9, p. 237-243]:

(1′) X_k is the *multitarget state*, i.e. the set of unknown target states (which are also of unknown number) and Z_k is the set of all measurements collected from all targets at time-step k;

(2′) $f_{k|k}(X_k|Z^{(k)})$ is a *multitarget posterior density* at time-set k conditioned on the time-stream $Z^{(k)} = \{Z_1, ..., Z_k\}$;

(3′) $f_k(Z|X)$ is the *multisensor, multitarget likelihood function* that describes the likelihood of observing the observation-set Z given that the multitarget system has multitarget state-set X;

(4′) $f_{k+1|k}(X_{k+1}|X_k)$ is the multitarget Markov transition density that describes the likelihood that the targets will have state-set X_{k+1} at time-step $k+1$ given that they had state-set at time-step k;

(5′) $f_{k+1|k}(X_{k+1}|Z^{(k)})$ is the time-prediction of the multitarget posterior $f_{k|k}(X_k|Z^{(k)})$ to time-step $k+1$; and

(6′) $f_{k+1}(Z_{k+1}|Z^{(k)}) = \int f_k(Z_{k+1}|Y)\, f_{k+1|k}(Y|Z^{(k)})\delta Y$ is the Bayes normalization constant.

(A short history of Bayes multitarget filtering can be found in section 3.6.) The multitarget filtering equations 29, 30 and 31 cannot be copied from the single-target filtering equations 1, 2, and 3 in the blind fashion just suggested [22, pp. 1-6, 91-93], [21], [25], [24]. For example,

(7′) the direct multitarget generalization of the MAP estimator is not defined in general (and, as previously indicated, the multitarget generalization of the posterior expectation still resists definition); $\hat{X}^{JoM}_{k+1|k+1}$ is a specially-defined multitarget analog of the MAP estimator [22, 40-44];

(8′) the integrals $\int \cdot\delta X_k$ and $\int \cdot\delta Y$ are set integrals.

Clearly, if the single-target Bayes filter equations 1, 2, and 3 are already computationally daunting, then the multitarget filter equations 29, 30, and 31 will be computationally intractable in most circumstances. So, we proceed by analogy from the single-target case. There, we assumed that SNR is so high that the posterior is approximately completely characterized by its first moment statistic and then propagated $\hat{\mathbf{x}}_{k|k}$ alone using a constant-gain Kalman filter. The idea underlying the PHD filter is to extend this reasoning to the multitarget case. We assume that SNR is so high that the first-order moment (the PHD) of the multitarget system is an approximate sufficient statistic: $f_{k|k}(X_k|Z^{(k)}) \cong f_{k|k}(X_k|\hat{D}_{k|k})$. We then "fill in the question marks" in the following diagram,

$$\begin{array}{ccccccc}
\cdots \rightarrow & f_{k|k} & \rightarrow^{\text{multitarget prediction}} & f_{k+1|k} & \rightarrow^{\text{multitarget Bayes' rule}} & f_{k+1|k+1} & \rightarrow \cdots \\
 & \downarrow & & \downarrow & & \downarrow & \\
\cdots \rightarrow & \hat{D}_{k|k} & \rightarrow^{\text{moment prediction??}} & \hat{D}_{k+1|k} & \rightarrow^{\text{moment data update??}} & \hat{D}_{k+1|k+1} & \rightarrow \cdots
\end{array}$$

where the top row portrays the time-evolution of the multitarget Bayes filtering equations 29 and 30; and where the downward-pointing arrows indicate the replacement of multitarget posteriors by their corresponding PHD's. Ideally, this diagram would be *Bayes-closed* in the sense of Kulhavý [12] and Iltis [11], meaning that for any k:

1. the two time-update paths $f_{k|k} \to f_{k+1|k} \to \hat{D}_{k+1|k}$ and $f_{k|k} \to \hat{D}_{k|k} \to \hat{D}_{k+1|k}$ yield the same $\hat{D}_{k+1|k}$;

2. the two data-update paths $f_{k+1|k} \to f_{k+1|k+1} \to \hat{D}_{k+1|k+1}$ and $f_{k+1|k} \to \hat{D}_{k+1|k} \to \hat{D}_{k+1|k+1}$ yield the same $\hat{D}_{k+1|k+1}$; and

3. $\hat{D}_{k+1|k}$ is a "best-fit approximation" of $f_{k+1|k}$, and $\hat{D}_{k+1|k+1}$ is a "best-fit approximation" of $f_{k+1|k+1}$, in some explicitly specified sense.

If all three of these properties were satisfied, the filter indicated in the bottom row of the diagram would be *theoretically guaranteed to not diverge.* That is, it would always produce exactly the same best-fit $\hat{D}_{k|k}$ or $\hat{D}_{k+1|k}$ that one would get if one could implement the optimal filter in the top row and then compress the respective multitarget posteriors into their corresponding first moments. In the approach proposed in this paper, property (2) is only approximately satisfied. The PHD filter is based on the following assumption: *the random multitarget track-set $\Gamma_{k|k}$ is approximately equal to the Poisson process (Poisson track-cluster) $\hat{\Gamma}_{k|k}$ that most closely resembles it.* By definition, Γ is a multidimensional (spatial) Poisson process with intensity function $D(\mathbf{x})$ if [48, p. 33], [43, pp. 86-87], [15, p. 6]:

1. for all $i > 1$, $\Gamma \cap S_1, ..., \Gamma \cap S_i$ are statistically independent random sets whenever $S_1, ..., S_i$ are mutually disjoint bounded subsets; and

2. the number of elements of Γ in any bounded subset S is Poisson distributed: $\Pr(|\Gamma \cap S| = n) = \frac{e^{-D(S)} D(S)^n}{n!}$, where $D(S) = \int_S D(\mathbf{x}) d\mathbf{x}$ is called the intensity measure.

The belief-mass function of a Poisson process has the form $\beta(S) = \exp(D(S) - N)$ where $N = \int D(\mathbf{x}) d\mathbf{x}$, the expected number of objects, is given by equation 28. The multi-object moment density function and multi-object posterior of a Poisson track-set $\Gamma_{k|k}$ have the simple forms [6, p. 226]:

$$\hat{m}_{k|k}(\emptyset) = 1, \qquad \hat{m}_{k|k}(X) = \prod_{\mathbf{x} \in X} \hat{D}_{k|k}(\mathbf{x}) \tag{32}$$

$$\hat{f}_{k|k}(\emptyset) = e^{-N_{k|k}}, \qquad \hat{f}_{k|k}(X) = e^{-\hat{N}_{k|k}} \prod_{\mathbf{x} \in X} \hat{D}_{k|k}(\mathbf{x}) \tag{33}$$

Equation 33 tells us that the joint density of the tracks $\mathbf{x}_1, ..., \mathbf{x}_n$, given that there are n tracks, is

$$\hat{f}_{k|k}(\mathbf{x}_1, ..., \mathbf{x}_n | n, Z^{(k)}) = \frac{1}{n!} \hat{f}_{k|k}(\{\mathbf{x}_1, ..., \mathbf{x}_n\} | n, Z^{(k)}) = \frac{\hat{f}_{k|k}(\{\mathbf{x}_1, ..., \mathbf{x}_n\} | Z^{(k)})}{\Pr(|\Gamma_{k|k}| = n)} = \prod_{i=1} \hat{s}_{k|k}(\mathbf{x}_i)$$

where $\hat{f}_{k|k}(\mathbf{x}_1, ..., \mathbf{x}_n | n, Z^{(k)})$ is the joint posterior distribution of the vector $(\mathbf{x}_1, ..., \mathbf{x}_n)$ and where $\hat{s}_{k|k}(\mathbf{x}) = \hat{D}_{k|k}(\mathbf{x}) / \hat{N}_{k|k}$ is the spatial probability distribution of each track [43, p. 90]. That is, given the number of tracks, their locations are independent, identically-distributed random vectors.[‖]

We now choose $\hat{f}_{k|k}$ so that it is the *multitarget distribution of the Poisson-distributed track-cluster $\hat{\Gamma}_{k|k}$ that best approximates the posterior track-set* $\Gamma_{k|k}$, in the sense that it minimizes the multitarget Kullback-Leibler discrimination [9, pp. 205-215]:

$$K(f_{k|k}; \hat{f}_{k|k}) = \int f_{k|k}(X | Z^{(k)}) \log \left(\frac{f_{k|k}(X | Z^{(k)})}{\hat{f}_{k|k}(X)} \right) \delta X \tag{34}$$

[‖]This fact tells us how to construct Poisson track clusters. Let $\mathbf{X}_1, \mathbf{X}_2, ...$ be an infinite sequence of independent, identically distributed random track-vectors and let ν be a Poisson-distributed random integer: $\Pr(\nu = n) = e^{-\hat{N}_{k|k}} \hat{N}^n_{k|k} / n!$. Then $\Gamma = \{\mathbf{X}_1, ..., \mathbf{X}_\nu\}$ is a Poisson track cluster, where it is assumed that $\Gamma = \emptyset$ whenever $\nu = 0$.

Minimization occurs (see section 5.3) when $\hat{D}_{k|k}(\mathbf{x}) = D_{k|k}(\mathbf{x}|Z^{(k)})$. This in turn implies that $\hat{N}_{k|k} = N_{k|k}$ where $N_{k|k} = \int D_{k|k}(\mathbf{x}|Z^{(k)})d\mathbf{x}$ is the expected number of tracks at time-step k. Consequently: when we say that the first-order moment $\hat{D}_{k|k}$ can be propagated in place of the multitarget posterior $f_{k|k}$, what we are actually saying is that we can propagate $\hat{f}_{k|k}$ in place of $f_{k|k}$ because it is a reasonably good approximation in the sense that the track-set $\Gamma_{k|k}$ strongly resembles the Poisson track-cluster $\hat{\Gamma}_{k|k}$: i.e., $\hat{f}_{k|k} \cong f_{k|k}$.

3.3. The Bayes First-Moment Filtering Equations

Given these preliminaries, it can be shown [17, Theorem 4], [27] that between measurement-collection times, the PHD can be propagated in time, *without approximation*, using the exact prediction integral

$$\hat{D}_{k+1|k}(\mathbf{y}|Z^{(k)}) = \int \left(d_{k+1|k}(\mathbf{x})f_{k+1|k}(\mathbf{y}|\mathbf{x}) + \hat{b}_{k+1|k}(\mathbf{y}|\mathbf{x})\right)\ \hat{D}_{k|k}(\mathbf{x}|Z^{(k)})d\mathbf{x} \tag{35}$$

where

(1″) $1 - d_{k+1|k}(\mathbf{x})$ is the probability that a target with state $\mathbf{x}$ at time-step k will disappear from the scene at time-step $k+1$;

(2″) $\hat{b}_{k+1|k}(\mathbf{y}|\mathbf{x})$ is the PHD of the multitarget density $b_{k+1|k}(X|\mathbf{x})$ that describes the likelihood that a target with state $\mathbf{x}$ at time-step k will generate a set X of new targets at time-step $k+1$;

If the track-set strongly resembles a Poisson track-cluster then it can be shown [17, Theorem 5], [27] that the following approximate equation is the Bayes update of the PHD using a new scan Z_{k+1} of data:

$$\begin{aligned}\hat{D}_{k+1|k+1}(\mathbf{x}|Z^{(k+1)}) \quad &\cong \quad \sum_{\mathbf{z}\in Z_{k+1}} \frac{p_D\hat{D}_{k+1}(\mathbf{z}|Z^{(k)})}{\lambda_{k+1}c_{k+1}(\mathbf{z}) + p_D\hat{D}_{k+1}(\mathbf{z}|Z^{(k)})}\hat{D}_{k+1|k+1}(\mathbf{x}|\mathbf{z}, Z^{(k)}) \\ &\quad + \frac{1-p_D}{1-(1-p_D)N_{k+1|k}}\hat{D}_{k+1|k}(\mathbf{x}|Z^{(k)})\end{aligned} \tag{36}$$

where:

(3″) $N_{k+1|k} = \int \hat{D}_{k+1|k}(\mathbf{x}|Z^{(k)})d\mathbf{x}$ is the predicted expected number of targets at time-step $k+1$;

(4″) p_D is the (state-independent) probability of detection of the sensor, assumed to be large enough that $p_D > 1 - N_{k+1|k}^{-1}$ for any k;

(5″) λ_{k+1} is the average number of Poisson false alarms per data-scan, and $c_{k+1}(\mathbf{z})$ is the (state-independent) distribution of each of these false alarms;

(6″) $\hat{D}_{k+1}(\mathbf{z}|Z^{(k)}) = \int f(\mathbf{z}|\mathbf{x})\ \hat{D}_{k+1|k}(\mathbf{x}|Z^{(k)})d\mathbf{x}$; and

(7″) $\hat{D}_{k+1|k+1}(\mathbf{x}|\mathbf{z}, Z^{(k)}) = f(\mathbf{z}|\mathbf{x})\ \hat{D}_{k+1|k}(\mathbf{x}|Z^{(k)})\ \hat{D}_{k+1}(\mathbf{z}|Z^{(k)})^{-1}$ is a Bayes' rule-like update of $\hat{D}_{k+1|k}(\mathbf{x}|Z^{(k)})$ using the observation $\mathbf{z}$.

It can also be shown [17, Theorem 6] that equation 36 is easily extended to deal with multiple sensors, assuming that the observation-sets collected by these sensors are conditionally independent.

In other words: if SNR is large enough then multitarget detection, tracking, and identification can be accomplished using a process that strongly resembles single-target nonlinear filtering—*and that also does not require report-to-track association.* Rather, data association is essentially replaced by *multi-peak extraction.* At each stage, the PHD filter propagates not only the PHD $\hat{D}_{k|k}(\mathbf{x}|Z^{(k)})$ but also the expected number of targets $\hat{N}_{k|k} = \int \hat{D}_{k|k}(\mathbf{x}|Z^{(k)})d\mathbf{x}$. Consequently, estimation of the multitarget state is accomplished by computing the nearest integer $[N_{k|k}]$ in $N_{k|k}$ and then searching for the $[N_{k|k}]$ largest peaks of $\hat{D}_{k|k}(\mathbf{x}|Z^{(k)})$.

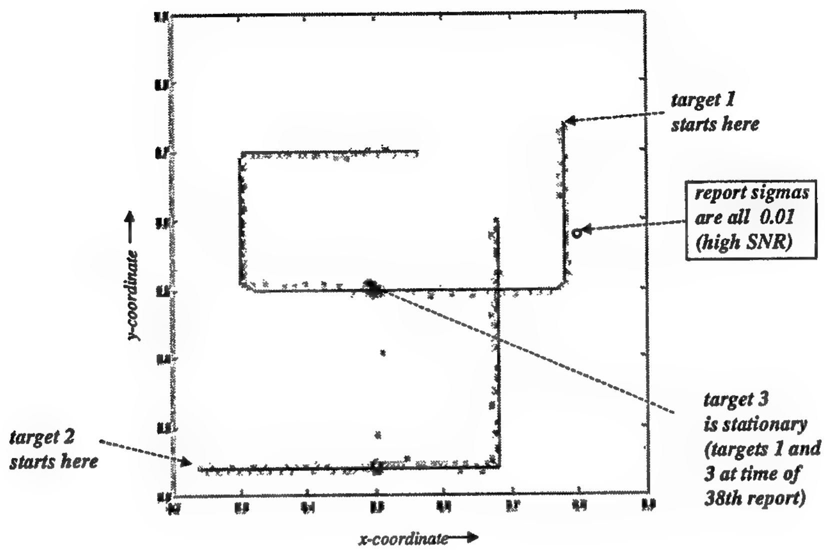

Figure 1. Simple Test Scenario. Target 1 enters at upper right of the scene, Target 2 enters at lower left, and Target 3 is near the center and motionless. Target 1 overruns Target 3 at the time of 38^{th} observation.

3.4. PHD Filter: Preliminary Results

The purpose of this section is to describe the performance of a preliminary implementation of the PHD filter, as applied to a simple model problem. This work is described more fully in recent conference papers [7], [29]. The PHD filtering equations 35 and 36 were implemented using a "spectral compression" computational nonlinear filtering technique developed at Lockheed Martin Tactical Systems. A simple multi-peak extraction algorithm was used to find the N largest peaks of the PHD at each time-step, where N is the integer-valued expected number of targets supplied by the PHD filter. This peak extractor divides the scenario into a quantized grid and searches for maximal values. Localization accuracy is thus limited by the quantization error.

The PHD filter was applied to the simple three-target model scenario pictured in Figure 1. Two moving targets enter the scene from upper right and lower left, whereas a third target is motionless at the center. One target overruns the motionless target at the time of the 38^{th} observation, generating confusion about target location. The scene is observed by a single Gaussian sensor. In keeping with the assumptions that underlie the PHD filter, signal-to-noise ratio is assumed to be high: $\sigma = .01$. There are no false alarms, and target number is assumed to be unknown but unchanging. A simple instantaneous straight-line motion model is used. With these assumptions, the PHD filtering equations 35 and 36 reduce to the simple form:

$$\hat{D}_{k+1|k}(\mathbf{y}|Z^{(k)}) = \int f_{k+1|k}(\mathbf{y}|\mathbf{x})\; \hat{D}_{k|k}(\mathbf{x}|Z^{(k)})d\mathbf{x}, \qquad \hat{D}_{k+1|k+1}(\mathbf{x}|Z^{(k+1)}) \cong \sum_{\mathbf{z}\in Z_{k+1}} \hat{D}_{k+1|k+1}(\mathbf{x}|\mathbf{z}, Z^{(k)})$$

Our results are as follows.

Figure 2 shows the graph of the PHD at the time of the 35^{th} observation. Despite the fact that some confusion between Target 1 and Target 3 is beginning to occur, the peaks corresponding to the three targets are clearly separated.

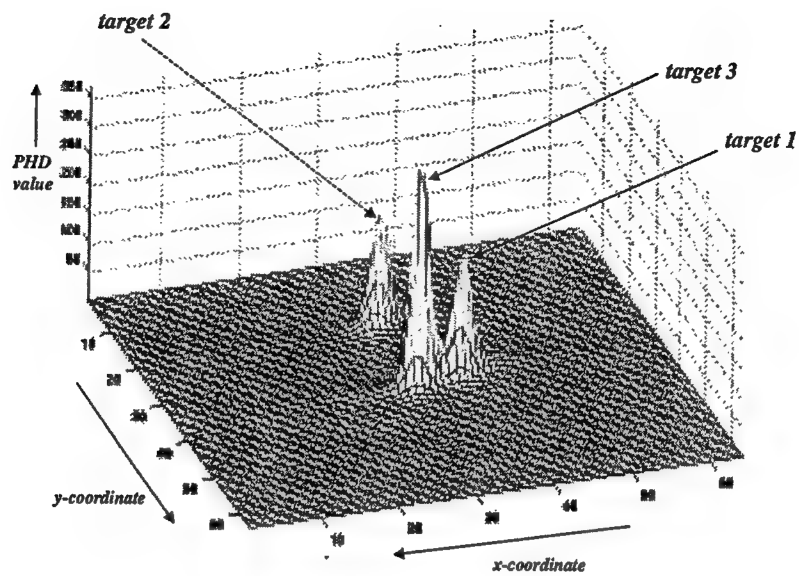

Figure 2. PHD at 35^{th} Time-Step. Since Target 3 is motionless, the peak corresponding to its location is large compared to the peaks for the other two targets.

Because Target 3 is motionless there is less confusion regarding its position, and so its peak is higher than the peaks corresponding to the other two targets.

Figure 3 shows the tracking performance of the PHD filter throughout the scenario, using a measure of performance called the "Hausdorff multitarget miss distance" [52], [22, p. 46], [9, pp. 137-138]. The Hausdorff distance $d_{Haus}(G,X)$ is the multitarget analog of the single-target concept of Euclidean miss distance $||\mathbf{g}-\mathbf{x}||$ between a ground-truth target $\mathbf{g}$ and a track estimate $\mathbf{x}$. It is defined as:

$$d_{Haus}(G,X) = \max\{d_0(G,X),\ d_0(X,G)\}, \qquad d_0(G,X) = \max_{\mathbf{g}\in G}\min_{\mathbf{x}\in X} ||\mathbf{g}-\mathbf{x}||$$

where G is the set of ground truth targets and X a the set of track estimates. In our case $G=\{\mathbf{g}_1,\mathbf{g}_2,\mathbf{g}_3\}$ is the indicated ground truth at a particular observation-time, and $X=\{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3\}$ are the track locations generated by the PHD via peak extraction, at the same observation-time.

Figure 3 shows that, over-all, the PHD filter successfully tracks the targets *even though it does not try to associate reports with tracks*. Tracking accuracy is limited by the quantization error of the peak-extraction algorithm rather than by the PHD, and is approximately $\sigma = 0.005$. It is good even before and after the time of the 38^{th} time-step, the time of maximal target-to-target confusion. The large error spikes at the 35^{th} and 40^{th} observations are caused by failures in the peak-extraction algorithm and not the PHD filter itself. (Figure 2 shows, for example, that the peak extractor failed at the 35^{th} observation even though the three peaks are clearly separated.) In these cases, the peak extractor failed to find one of the peaks and assigned two target locations to another peak.

This experiment demonstrates that the development of good multi-peak extraction algorithms is critical to the successful implementation of a PHD filter. It also shows that, to a great extent, peak extraction plays the same role in a PHD filter that report-to-track association does in a conventional multi-hypothesis multitarget tracker. The

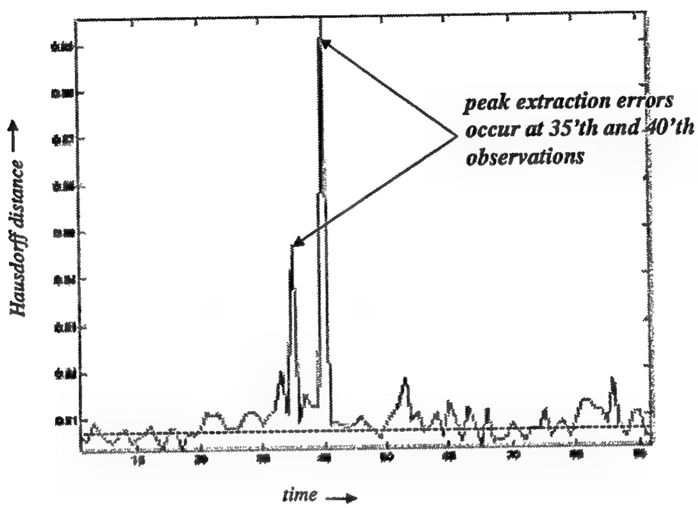

Figure 3. Tracking Performance of PHD Filter. The Hausdorff multitarget miss distance (vertical axis) is plotted versus observation number. Localization accuracy is good even when Target 1 overruns Target 3 at the time of the 38^{th} observation. The two large peaks at the 35^{th} and 40^{th} time-steps are due to peak-extraction failures.

computational complexity involved in data association is not entirely sidestepped, but rather is partially transferred to the peak-extraction process.

3.5. Open Research Question: Expected Track-Sets

Many definitions of the set-valued expectation $E[\Gamma]$ of a random set Γ have been proposed [9, pp. 176-177], but none of these are applicable to random finite sets—i.e., to simple point processes. At the 1998 GTRI/ONR Workshop on Tracking and Sensor Fusion [28] I suggested a definition of a track-valued multitarget expected value of a random track-set, based on the idea of transforming finite sets into multinomials (similar to a similar transformation proposed earlier [9, p. 179]). This definition of $E[\Gamma]$ produces intuitively acceptable multitarget expectations in special cases. For example, if the track-set $\Gamma = \{\mathbf{X}_1, ..., \mathbf{X}_n\}$ is the union of statistically independent tracks $\mathbf{X}_1, ..., \mathbf{X}_n$ whose respective expected values are $\bar{\mathbf{X}}_1, ..., \bar{\mathbf{X}}_n$, then $E[\Gamma] = \{\bar{\mathbf{X}}_1, ..., \bar{\mathbf{X}}_n\}$. However, the approach does not appear to be satisfactory in general. The problem of correctly defining $E[\Gamma]$ is an open research question. In this section I argue (as I did in the 1998 workshop) that any acceptable definition of $E[\Gamma]$ should have the following properties:

- *Consistency:* if $\mathbf{X}$ is any random vector then

$$E[\{\mathbf{X}\}] = \{E[\mathbf{X}]\}$$

  That is, the expectation of a random track-set that always contains a single random track should itself contain only a single track; and this track should be the expected value of the random track: $\overline{\{\mathbf{X}\}} = \{\bar{\mathbf{X}}\}$

- *Empty track-sets are ignored*: Let Γ be a random track-set and let $E[\Gamma|\Gamma \neq \emptyset]$ denote the conditional expectation of Γ given that it is non-empty. (That is, whereas $E[\Gamma]$ is the expectation with respect to the

multitarget density $f_\Gamma(X)$, $\mathrm{E}[\Gamma|\Gamma \neq \emptyset]$ is the expectation with respect to the multitarget density $f_{\Gamma|\Gamma\neq\emptyset}$ defined by $f_{\Gamma|\Gamma\neq\emptyset}(X) = (1 - f_\Gamma(\emptyset))^{-1} f_\Gamma(X)$ when $X \neq \emptyset$ and $f_{\Gamma|\Gamma\neq\emptyset}(\emptyset) = 0$ otherwise.) Then

$$\mathrm{E}[\Gamma] = \mathrm{E}[\Gamma|\Gamma \neq \emptyset]$$

That is, empty track-sets should provide no contribution to the value of a multitarget expected value.

- *Preservation of geometry*: Let Γ_1, Γ_2 be *statistically independent* track-sets. Then

$$\mathrm{E}[\Gamma_1 \cup \Gamma_2] = \mathrm{E}[\Gamma_1] \cup \mathrm{E}[\Gamma_2]$$

That is, when two random track-clusters are statistically non-interacting, their joint multitarget expected value should just be the union of the multitarget expected values of the individual track-clusters. For example, $\overline{\{\mathbf{X}_1, \mathbf{X}_2\}} = \{\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2\}$ if $\mathbf{X}_1, \mathbf{X}_2$ are independent random vectors.

- *Preservation of affine transformations*: Let Γ be a random track-set and $\mathbf{a}$ a constant track-vector and $T(\mathbf{x})$ any linear transformation of track-vectors $\mathbf{x}$. Define $T(\{\mathbf{x}_1, ..., \mathbf{x}_n\}) = \{T\mathbf{x}_1, ..., T\mathbf{x}_n\}$ and $X + \{\mathbf{a}\} = \{\mathbf{x} + \mathbf{a} | \ \mathbf{x} \in X\}$ for any subset X. Then

$$\mathrm{E}[T(\Gamma) + \{\mathbf{a}\}] = T(\mathrm{E}[\Gamma]) + \{\mathbf{a}\}$$

That is, to compute the multitarget expected value of a random track-set after a (possibly degenerate) affine transformation $\mathbf{x} \longmapsto T(\mathbf{x}) + \mathbf{a}$ of the coordinate system, just apply the same transformation to its expected value. For example, $\overline{\{a \cdot \mathbf{X}_1 + \mathbf{a},\ a \cdot \mathbf{X}_2 + \mathbf{a}\}} = a \cdot \overline{\{\mathbf{X}_1, \mathbf{X}_2\}} + \{\mathbf{a}\}$ where $\mathbf{X}_1, \mathbf{X}_2$ are (not necessarily independent) random vectors and a is a constant.

3.6. Related Approaches

The idea of using a single-target density function $g_{k|k}$ (or the probability contours of its graph) as a basis for multitarget tracking is a relatively common one. Examples of implemented algorithms are the Naval Research Laboratory's TABS (Tactical Antisubmarine-warfare Battle-management System) tracker, Metron Corp.'s *Nodestar* tracker [47] and the "probabilistic mapping" multitarget tracking approach of Tao, Abileah, and Lawrence [49]. The PHD filter differs from this work in that it (1) provides a solid theoretical foundation for single-density approaches in general, (2) clarifies the relationship between single-density approaches and the optimal multitarget filter, (3) makes explicit the theoretical assumptions that implicitly underlie single-density approaches (in particular, the assumption of high SNR), and (4) validates Stein and Winter's intuition that the PHD is a "theoretically correct" choice of $g_{k|k}$ as well as their concept of "weak evidence accrual." Beyond this, the PHD approach leads to new practical insight and new implementations because the PHD can be propagated by conventional nonlinear filtering equations.

The concept of *multitarget Bayesian nonlinear filtering* (equations 29 and 30) is a relatively new one. If one assumes that the number of targets is known beforehand, the earliest exposition appears to be due to Washburn [51] in 1987, who used a random measure-based point process approach. When the number n of targets is *not* known and must be determined along with the individual target states, the earliest work appears to be due to Miller, O'Sullivan, Srivastava, et. al. [44], [45]. Their very sophisticated approach utilizes solution of stochastic diffusion equations on non-Euclidean manifolds. It is also apparently the only approach to deal with continuous evolution of the multitarget state. Mahler was apparently the first to systematically deal with the general discrete state-evolution case (Bethel and Paras [3] assume discrete observation and state variables). Stone et. al. have provided a valuable contribution by clarifying the relationship between multitarget Bayes filtering and multi-hypothesis correlation [47, pp. 161-208], [9, p. 32], [22, p. 48]. Nevertheless their approach is, with regrets, best described as "heuristic" for reasons described more fully elsewhere [22, pp. 91-93]. Kastella's "joint multitarget probabilities (JMP)," introduced at Lockheed Martin Tactical Systems in 1996, are a renaming of a number of early core FISST concepts

(set integrals, multitarget information metrics, multitarget posteriors, joint multitarget state estimators, etc.) devised two years earlier [37, pp. 27-28], [25]. Portenko et. al., also using a random measure-based point process approach, use branching-process concepts to model target appearance and disappearance [38].

Most recently, M. Kouritzin has noted [16] that the theory underlying particle-systems filters also subsumes point processes and therefore multitarget tracking as well. In particular, this means that particle system-based multitarget nonlinear filters will have very good convergence properties (i.e., for every multitarget observation sequence, the particle density will converge almost surely to the actual multitarget distribution). It also means that particle-system methods can be applied to single- and multitarget systems whose motion models—e.g., heavy-tail or non-differentiable models—are too ill-behaved to be describable by conventional Kolmogorov-Forward (Fokker-Planck) equations.

It should also be pointed out that Mori, Chong, et. al. first proposed random set theory as a potential foundation for multitarget detection, tracking, and identification (although within a multi-hypothesis framework) [35, pp. 33-37]. Since 1995 Mori has returned to the field and published a number of very interesting papers based on random set ideas [32], [33], [34].

4. SECOND-ORDER MULTITARGET STATISTICS

The PHD filter described in section 3 makes use only of the first-order multitarget moment statistic and, in this sense, is a statistical analog of single-target, constant-gain Kalman filters such as the α-β-γ filter. The Kalman filter arises when we try to propagate the second-order as well as first-order statistics of a target. The extended Kalman filter (EKF) results when we expand the likelihood functions and Markov densities in Taylor's series, in order to propagate second-order approximations of the single-target posterior density function. It is natural to ask whether analogous things can be done for multitarget systems. First, can we propagate a second-order approximation of the multitarget posterior density (i.e., the PHD and a multitarget covariance) instead of the multitarget posterior itself? If so, we would have a statistical multitarget analog of the Kalman filter. Second, can we expand the multitarget likelihood function and multitarget Markov density in some kind of "multitarget Taylor's series"? If so, we would have a leg up on defining a direct multitarget analog of the EKF. The purpose of this section is to discuss these issues. Our results are preliminary, but can be summarized as follows. At this time, it does not appear likely that a computationally tractable "multitarget Kalman filter" can be constructed from second-order multitarget moment statistics. However, it appears that it might be possible to develop a multitarget analog of the EKF, though the potential computational practicality of a "multitarget EKF" remains to be seen.

The section is organized as follows. In section 4.1 I introduce the multitarget analog of covariance-moment statistics, the *multitarget covariance density function.* Then, in section 4.2, I discuss the possibility of constructing multitarget filters based on second-order multitarget covariance statistics—i.e., statistical multitarget analogs of the Kalman filter. In section 4.3 I revisit the EKF and identify a "missing link"—namely, multitarget Taylor's series expansions—that must be dealt with before any multitarget analog of the EKF can be constructed. Finally, in section 4.4, I introduce a generalization of the concept of a point target—the "point-Poisson track-cluster"—and show how it leads to Taylor's series expansions of the multitarget log-likelihood.

4.1. The Multitarget Covariance Density Function

In section 2.1.6 I introduced the factorial cumulant densities of a point process. Applying equation 19 to the random track-set $\Gamma_{k|k}$, we bundle the factorial cumulant densities $c_{\Gamma_{k|k},[n]}(\mathbf{x}_1,...,\mathbf{x}_n)$ into a single density function defined on a finite-set variable $X=\{\mathbf{x}_1,...,\mathbf{x}_n\}$ and call it the *multitarget covariance density function* of the track-set:

$$c_{k|k}(\emptyset|Z^{(k)})=0, \qquad c_{k|k}(X|Z^{(k)})=c_{\Gamma_{k|k},[n]}(\mathbf{x}_1,...,\mathbf{x}_n)$$

From equation 19 it follows that

$$c_{k|k}(X|Z^{(k)}) = \frac{\delta \log \beta_{k|k}}{\delta X}(S|Z^{(k)})\bigg|_{S=\mathbb{S}} = \frac{\delta \log \beta_{k|k}}{\delta X}(\mathbb{S}|Z^{(k)})$$

where $\mathbb{S}$ is the entire state space. Thus, the multitarget covariance density can, like the multitarget posterior and moment density functions, be computed using iterated set derivatives. Likewise, equation 20 becomes

$$\log G_{k|k}[1+h] = \int h^X \cdot c_{k|k}(X|Z^{(k)})\delta X$$

where as usual $h^X = \prod_{\mathbf{x}\in X} h(\mathbf{x})$ given $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$. As before, we note that $c_{k|k}(\{\mathbf{x}\}|Z^{(k)}) = m_{k|k}(\{\mathbf{x}\}|Z^{(k)}) = D_{k|k}(\mathbf{x}|Z^{(k)})$ and we give the second-order covariance

$$c_{k|k}(\mathbf{x}_1, \mathbf{x}_2|Z^{(k)}) = c_{k|k}(\{\mathbf{x}_1, \mathbf{x}_2\}|Z^{(k)})$$

a special name: it is the *covariance density* of the random track-set $\Gamma_{k|k}$.

As an example, if $\Gamma_{k|k}$ is a Poisson track-cluster then $c_{k|k}(X|Z^{(k)}) = 0$ for all $|X| > 1$ [6, p. 226].

4.2. Second-Order Filtering

The PHD filter of section 3.3 was based on the assumption that the PHD is an approximate sufficient statistics for the multitarget posterior $f_{k|k}(X|Z^{(k)})$. In other words, the track-set $\Gamma_{k|k}$ is assumed to be approximately Poisson and thus the multitarget posterior is approximately the multitarget density function of a Poisson process. Stated in the language of multitarget covariances, this is the same thing as insisting that $c_{k|k}(X|Z^{(k)}) \cong 0$ for all $|X| > 1$. Suppose that we instead insist that $c_{k|k}(X|Z^{(k)}) = 0$ for all $|X| > 2$. This means that $\Gamma_{k|k}$ would be the formal point process statistical analog of a Gaussian random vector. Such a track-cluster is known in point process theory as a "Gauss-Poisson process" [6, 266-267]. Any Gauss-Poisson process is essentially just a Poisson process in which each track is replaced by a pair of correlated tracks [6, pp. 247-248].

If we were to follow the strategy used in the derivation of the PHD filtering equations 35 and 36, we would proceed as follows. We would assume that $f_{k|k}(X|Z^{(k)})$ is Gauss-Poisson and that, consequently, it can be written as a combinatorial sum of products of $c_{k|k}(\mathbf{x}_1, \mathbf{x}_2|Z^{(k)})$ and $D_{k|k}(\mathbf{x}|Z^{(k)})$. Then we would apply the multitarget Markov prediction integral (equation 29) to derive $f_{k+1|k}(X|Z^{(k)})$ and then compute the first-moment density $D_{k+1|k}(\mathbf{x}|Z^{(k)})$ and covariance density $c_{k+1|k}(\mathbf{x}_1, \mathbf{x}_2|Z^{(k)})$. We already know how to determine $D_{k+1|k}(\mathbf{x}|Z^{(k)})$ exactly from equation 35. It turns out that $c_{k+1|k}(\mathbf{x}_1, \mathbf{x}_2|Z^{(k)})$ can be computed by applying this same equation to both arguments of $c_{k|k}(\mathbf{x}_1, \mathbf{x}_2|Z^{(k)})$. The difficulty comes with the multitarget Bayes' rule, equation 30. We would apply the multitarget likelihood and Bayes' rule to get $f_{k+1|k+1}(X|Z^{(k+1)})$ and then try to compute $D_{k+1|k+1}(\mathbf{x}|Z^{(k+1)})$ and $c_{k+1|k+1}(\mathbf{x}_1, \mathbf{x}_2|Z^{(k+1)})$. However, the answer to this symbolic computation currently eludes solution. Moreover, even if it were completed successfully the approximate Bayes' rule for propagating $c_{k+1|k}(\mathbf{x}_1, \mathbf{x}_2|Z^{(k)})$ to $c_{k+1|k+1}(\mathbf{x}_1, \mathbf{x}_2|Z^{(k+1)})$ may be so complex as to be unusable in practice.

Consequently another approach is called for, as suggested in the next section.

4.3. Basic Idea for a Multitarget EKF

The usual development of the extended Kalman filter [5, p. 109-111] begins with a sensor measurement and target motion model

$$\mathbf{z}_k = \mathbf{g}_k(\mathbf{x}_k) + \mathbf{V}_k, \qquad \mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{W}_k$$

where $\mathbf{f}_k$ and $\mathbf{g}_k$ are nonlinear vector transformations and $\mathbf{V}_k, \mathbf{W}_k$ are Gaussian random noise processes. These models are then linearized:

$$\begin{aligned} \mathbf{g}_k(\mathbf{x}) &\cong \mathbf{g}_k(\hat{\mathbf{x}}_{k|k-1}) + A_k\left(\mathbf{x} - \hat{\mathbf{x}}_{k|k-1}\right), \qquad \mathbf{f}_k(\mathbf{x}) \cong \mathbf{f}_k(\hat{\mathbf{x}}_{k|k}) + B_k\left(\mathbf{x} - \hat{\mathbf{x}}_{k|k}\right) \\ A_k &= \frac{\partial \mathbf{g}_k}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k|k-1}), \qquad B_k = \frac{\partial \mathbf{f}_k}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k|k}) \end{aligned}$$

where $\frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\hat{\mathbf{x}}) = \left(\frac{\partial h_i}{\partial x_j}(\hat{\mathbf{x}})\right)_{i,j}$ denotes the Jacobian matrix of the vector transformation $\mathbf{h}$, evaluated at $\mathbf{x} = \hat{\mathbf{x}}$.

It is not possible to directly extend this line of reasoning to the multitarget case. For example, suppose that two statistically independent targets with states $\mathbf{x}_1$ and $\mathbf{x}_2$ are present, and that in addition to the noise model $\mathbf{z}_k = \mathbf{g}_k(\mathbf{x}_k) + \mathbf{V}_k$ the sensor has no missed detections but its observations are corrupted by an independent Poisson false alarm process C_k. Then the typical multitarget observation would be:

$$Z_k = \{\mathbf{g}_k(\mathbf{x}_1) + \mathbf{V}_{k,1}\} \cup \{\mathbf{g}_k(\mathbf{x}_2) + \mathbf{V}_{k,2}\} \cup C_k$$

where $\mathbf{V}_{k,1}, \mathbf{V}_{k,2}$ are independent, identically distributed copies of $\mathbf{V}_k$. Because of independence, this can be converted into a measurement model on random densities:

$$\zeta_k = \delta_{\mathbf{g}_k(\mathbf{x}_1)+\mathbf{V}_{k,1}} + \delta_{\mathbf{g}_k(\mathbf{x}_2)+\mathbf{V}_{k,2}} + \delta_{C_k}$$

This is a multitarget analog of the equation $\mathbf{z}_k = \mathbf{g}_k(\mathbf{x}) + \mathbf{V}_k$, but it does not help us very much since it cannot be linearized in the multitarget state variable $\xi = \delta_{\mathbf{x}_1} + \delta_{\mathbf{x}_2}$.

If we are to generalize the EKF approximation strategy to the multitarget case, we must therefore approach it from a different direction. The above nonlinear-Gaussian models can be equivalently expressed in terms of a Markov density $f_{k+1|k}(\mathbf{y}|\mathbf{x})$ and a likelihood function $L_{\mathbf{z}_k}(\mathbf{x}) = f_k(\mathbf{z}_k|\mathbf{x})$ whose logarithms are

$$\begin{aligned} \log f_{k+1|k}(\mathbf{y}|\mathbf{x}) &= \log N_{R_k}(0) - \frac{1}{2}\left(\mathbf{y} - \mathbf{f}_k(\mathbf{x})\right)^T R_k^{-1}\left(\mathbf{y} - \mathbf{f}_k(\mathbf{x})\right) \\ \log L_{\mathbf{z}_k}(\mathbf{x}) &= \log N_{Q_k}(0) - \frac{1}{2}\left(\mathbf{z}_k - \mathbf{g}_k(\mathbf{x})\right)^T Q_k^{-1}\left(\mathbf{z}_k - \mathbf{g}_k(\mathbf{x})\right) \end{aligned}$$

Using the EKF approximation, the log-likelihood can be rewritten as

$$\begin{aligned} \log L_{\mathbf{z}_k}(\mathbf{x}) &\cong \log N_{Q_k}(0) - \frac{1}{2}\left(\begin{array}{c} \mathbf{z}_k - \mathbf{g}_k(\hat{\mathbf{x}}_{k|k-1}) \\ -A_k\left(\mathbf{x} - \hat{\mathbf{x}}_{k|k-1}\right) \end{array}\right)^T Q_k^{-1}\left(\begin{array}{c} \mathbf{z}_k - \mathbf{g}_k(\hat{\mathbf{x}}_{k|k-1}) \\ -A_k\left(\mathbf{x} - \hat{\mathbf{x}}_{k|k-1}\right) \end{array}\right) \\ &= \log L_{\mathbf{z}_k}(\hat{\mathbf{x}}_{k|k-1}) - \left(\mathbf{z}_k - \mathbf{g}_k(\hat{\mathbf{x}}_{k|k-1})\right)^T Q_k^{-1} A_k\left(\mathbf{x} - \hat{\mathbf{x}}_{k|k-1}\right) \\ &\quad -\frac{1}{2}\left(\mathbf{x} - \hat{\mathbf{x}}_{k|k-1}\right)^T A_k^T Q_k^{-1} A_k\left(\mathbf{x} - \hat{\mathbf{x}}_{k|k-1}\right) \end{aligned}$$

or, equivalently, as a truncated Taylor's series expansion:

$$\log L_{\mathbf{z}_k}(\mathbf{x}) \cong \log L_{\mathbf{z}_k}(\hat{\mathbf{x}}_{k|k-1}) + \frac{\partial \log L_{\mathbf{z}_k}}{\partial(\mathbf{x} - \hat{\mathbf{x}}_{k|k-1})}(\hat{\mathbf{x}}_{k|k-1}) + \frac{1}{2}\frac{\partial^2 \log L_{\mathbf{z}_k}}{\partial(\mathbf{x} - \hat{\mathbf{x}}_{k|k-1})^2}(\hat{\mathbf{x}}_{k|k-1}) \tag{37}$$

where

$$\frac{\partial \log L_{\mathbf{z}_k}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k|k-1}) = -\left(\mathbf{z}_k - \mathbf{g}_k(\hat{\mathbf{x}}_{k|k-1})\right)^T Q_k^{-1} A_k \mathbf{x}, \qquad \frac{\partial \log L_{\mathbf{z}_k}}{\partial \mathbf{x}_1 \partial \mathbf{x}_2}(\hat{\mathbf{x}}_{k|k-1}) = -\mathbf{x}_1^T A_k^T Q_k^{-1} A_k \mathbf{x}_2$$

(The logarithm of the Markov density can, of course, be expanded in the same way.)

In order to apply this line of reasoning to the multitarget case, we would need to be able to generalize equation 37 to a similar multitarget equation

$$\log L_{Z_k}(X) \cong \log L_{Z_k}(\hat{X}_{k|k-1}) + \frac{\partial \log L_{Z_k}}{\partial(X - \hat{X}_{k|k-1})}(\hat{X}_{k|k-1}) + \frac{1}{2}\frac{\partial^2 \log L_{Z_k}}{\partial(X - \hat{X}_{k|k-1})^2}(\hat{X}_{k|k-1}) \tag{38}$$

where Z_k is the multitarget measurement, X,$\hat{X}_{k|k-1}$ are multitarget states, and $L_{Z_k}(X) = f_k(Z_k|X)$ is the multitarget likelihood function. This cannot be accomplished because the arithmetic difference $X - \hat{X}_{k|k-1}$ is not defined, and because we do not know how to compute Frechét derivatives $\frac{\partial F}{\partial X}(Y)$ of functions $F(X)$ of a finite-set variable X. This, in turn, is due to the fact that $F(X)$ is "inherently non-differentiable" because of the *discrete jumps in target number* that occur in the argument X. In the next section, I propose the following remedy: extending $L_{Z_k}(X)$ to a *multitarget likelihood functional* $L_{Z_k}[h]$ that *is* differentiable in the function-valued argument h with respect to the Frechét functional derivative $(\partial F/\partial g)[h]$ defined in equation 10 of section 2.1.3.

4.4. Generalized Targets: Point Track-Clusters

We propose the following idea: extend the concept of a point target with state $\mathbf{x}$ to that of a *point track-cluster* with state $(a, \mathbf{x})$. By this I mean a cluster of targets, all of which are co-located at $\mathbf{x}$ and the expected number of which is $a > 0$. If $a = 1$ then the point cluster $(1, \mathbf{x})$ models a point target. A group of point clusters is just a finite set of the form $\mathcal{X} = \{(a_1, \mathbf{x}_1), ..., (a_n, \mathbf{x}_n)\}$. In section 2.1.1 we noted (equation 8) that $\mathcal{X}$ is mathematically equivalent to the density function $h = a_1\delta_{\mathbf{x}_1} + ... + a_n\delta_{\mathbf{x}_n}$. This identification allows us, in turn, to interpret *any* function $h(\mathbf{x})$ as a *continuously infinite collection of point target-clusters*—meaning that a point track-cluster is located at each point $\mathbf{x}$ of state space and the expected number-density of targets in this cluster is $h(\mathbf{x})$.

Given this, suppose that we have a single sensor with single-target likelihood function $L_{\mathbf{z}_k}(\mathbf{x}) = f_k(\mathbf{z}_k|\mathbf{x})$ and state-dependent probability of detection $p_D(\mathbf{x})$, which we assume can be extended to a function $p_D(a, \mathbf{x})$ of target number a as well as target state $\mathbf{x}$. Then I will show how to define a *likelihood functional* $L_{Z_k}[h] = f_k(Z_k|h)$ that extends the multitarget likelihood function $L_{Z_k}(X) = f_k(Z_k|X)$ and which describes the statistics of the observation-sets Z_k generated by all targets in the continuously infinite group h of point target-clusters. This then leads to the desired multitarget generalization of equation 37:

$$\log L_{Z_k}(X) \cong \log L_{Z_k}(\hat{X}_{k|k-1}) + \frac{\partial \log L_{Z_k}}{\partial(\delta_X - \delta_{\hat{X}_{k|k-1}})}[\delta_{\hat{X}_{k|k-1}}] + \frac{1}{2}\frac{\partial^2 \log L_{Z_k}}{\partial(\delta_X - \delta_{\hat{X}_{k|k-1}})^2}[\delta_{\hat{X}_{k|k-1}}]$$

where the sum on the right is a Taylor's series expansion in terms of the general functional derivatives (equation 11) introduced in section 2.1.3.

Note: For present purposes, I am not proposing the use of the multitarget likelihood functional $L_{Z_k}[h]$ in place of the multitarget likelihood function $L_{Z_k}(X)$. This would entail the replacement of multitarget density functions $f_{k|k}(X|Z^{(k)})$ by multitarget density functionals $f_{k|k}[h|Z^{(k)}]$ and the use of functional nonlinear filtering equations of the form

$$\begin{aligned} f_{k+1|k}[h|Z^{(k)}] &= \int f_{k+1|k}[h|g]\;\; f_{k|k}[g|Z^{(k)}]\mathcal{D}g \\ f_{k+1|k+1}[h|Z^{(k+1)}] &= \frac{L_{Z_{k+1}}[h]\;\; f_{k+1|k}[h|Z^{(k)}]}{f_k(Z_{k+1}|Z^{(h)})}, \qquad f_k(Z_{k+1}|Z^{(h)}) = \int L_{Z_{k+1}}[h]\;\; f_{k+1|k}[h|Z^{(k)}]\mathcal{D}h \end{aligned} \tag{39}$$

where $\int \cdot \mathcal{D}h$ is a suitable multi-dimensional generalization of theWiener *functional integral* [40, pp. 159, 186-188], [8], [31]. This is because we are searching for estimates that are true multitarget states $h = \delta_{\hat{X}_{k|k}} \Longleftrightarrow \hat{X}_{k|k}$ rather

than abstract functions (PHDs) $\hat{h}_{k|k}$. If we modeled multitarget systems using PHD's, on the other hand, then equations 39 would be the appropriate theoretical approach.

I also show that the likelihood functional can be approximated as a functional Gaussian distribution

$$L_Z(X) \cong L_Z[h_0] \cdot \exp\left(-\frac{1}{2}\int \frac{(\delta_X(\mathbf{y}) - h_0(\mathbf{y}))^2}{\sigma^2(\mathbf{y})} d\mathbf{y}\right)$$

if there is a unique function $h = h_0$ that solves the implicit differential equation

$$h(\mathbf{y})\frac{\partial p_D}{\partial a}(h(\mathbf{y}),\mathbf{y}) + p_D(h(\mathbf{y}),\mathbf{y}) = 0 \tag{40}$$

and if, in addition, the function

$$\frac{1}{\sigma^2(\mathbf{y})} = 2\frac{p_D(h_0(\mathbf{y}),\mathbf{y})}{h_0(\mathbf{y})^2} - \frac{\partial^2 p_D}{\partial a^2}(h_0(\mathbf{y}),\mathbf{y}) > 0 \tag{41}$$

is integrable and positive-valued for all $\mathbf{y}$.

In more detail: Suppose that we are given a sensor with single-target measurement model $\mathbf{z} = \mathbf{g}(\mathbf{x}) + \mathbf{W}$, likelihood function $f(\mathbf{z}|\mathbf{x}) = f_{\mathbf{W}}(\mathbf{z} - \mathbf{g}(\mathbf{x}))$, and state-dependent probability of detection $p_D(\mathbf{x})$. Suppose that Γ is a Poisson point track-cluster with intensity function $D(\mathbf{y})$ where $a = \int D(\mathbf{y})d\mathbf{y} > 0$ is the expected number of targets. Also, assume that probability of detection is a function $p_D(a, \mathbf{x})$ of target number a as well as target state $\mathbf{x}$, with $p_D(1, \mathbf{x}) = p_D(\mathbf{x})$. Then it is easily shown (see section 5.4) that the multitarget observation-set is also Poisson and that its multitarget likelihood function is

$$f(Z|a,\mathbf{x}) = e^{-a \cdot p_D(a,\mathbf{x})} \prod_{\mathbf{z}\in Z} (a \cdot p_D(a,\mathbf{x}) \cdot f(\mathbf{z}|\mathbf{x}))$$

Suppose, next, that we have a collection $\mathcal{X} = \{(a_1, \mathbf{x}_1), ..., (a_n, \mathbf{x}_n)\}$ of several independent point target-clusters. Then the multitarget likelihood function is again Poisson and is (see section 5.5):

$$f(Z|\mathcal{X}) = e^{-a_{\mathcal{X}}} \prod_{\mathbf{z}\in Z} D(\mathbf{z}|\mathcal{X}), \qquad D(\mathbf{z}|\mathcal{X}) = \sum_{i=1}^{n} a_i p_D(a_i, \mathbf{x}_i) f(\mathbf{z}|\mathbf{x}_i), \qquad a_{\mathcal{X}} = \sum_{i=1}^{n} a_i p_D(a_i, \mathbf{x}_i)$$

By analogy, suppose that the number n of point clusters increases continuously without bound in such a manner that the empirical distribution $\delta_{\mathbf{x}_1} + ... + a_n\delta_{\mathbf{x}_n}$ converges to some positive-valued function h. Then h can be interpreted as a continuously infinite family of point track-clusters, such that the point track-cluster located at $\mathbf{x}$ contains an average number (density) $h(\mathbf{x})$ of targets. The expected number of targets over the entire space is, therefore, $a_h = \int h(\mathbf{x})d\mathbf{x}$ if the integral exists. By analogy, say that the *likelihood functional* describing the observations generated by all targets is

$$L_Z[h] = f(Z|h) = e^{-\alpha[h]} \prod_{\mathbf{z}\in Z} D_{\mathbf{z}}[h]$$

where

$$D_{\mathbf{z}}[h] = \int h(\mathbf{x})\, p_D(h(\mathbf{x}),\mathbf{x})\, f(\mathbf{z}|\mathbf{x})d\mathbf{x}, \qquad \alpha[h] = \int h(\mathbf{x})\, p_D(h(\mathbf{x}),\mathbf{x})d\mathbf{x}$$

Notice that $\log L_Z[h]$ is continuous resp. differentiable in h if $p_D(a, \mathbf{x})$ is continuous resp. differentiable in a. In fact, application of the functional derivative (equation 10) and a little algebra shows (see section 5.6) that

$$\begin{aligned}
\frac{\delta D_{\mathbf{z}}}{\delta \mathbf{y}}[h] &= \left(h(\mathbf{y}) \frac{\partial p_D}{\partial a}(h(\mathbf{y}), \mathbf{y}) + p_D(h(\mathbf{y}), \mathbf{y}) \right) f(\mathbf{z}|\mathbf{y}) \\
\frac{\delta \alpha}{\delta \mathbf{y}}[h] &= h(\mathbf{y}) \frac{\partial p_D}{\partial a}(h(\mathbf{y}), \mathbf{y}) + p_D(h(\mathbf{y}), \mathbf{y}) \\
\frac{\delta \log L_Z}{\delta \mathbf{y}}[h] &= \left(-1 + \sum_{\mathbf{z} \in Z} \frac{f(\mathbf{z}|\mathbf{y})}{D_{\mathbf{z}}[h]} \right) \left(h(\mathbf{y}) \frac{\partial p_D}{\partial a}(h(\mathbf{y}), \mathbf{y}) + p_D(h(\mathbf{y}), \mathbf{y}) \right)
\end{aligned}$$

In like manner, a little more algebra shows that the second derivative of the log-likelihood is

$$\begin{aligned}
\frac{\delta^2 \log L_Z}{\delta \mathbf{y}_2 \delta \mathbf{y}_1}[h] &= -\left(\sum_{\mathbf{z} \in Z} \frac{f(\mathbf{z}|\mathbf{y}_1) f(\mathbf{z}|\mathbf{y}_2)}{D_{\mathbf{z}}[h]^2} \right) \left(h(\mathbf{y}_1) \frac{\partial p_D}{\partial a}(h(\mathbf{y}_1), \mathbf{y}_1) + p_D(h(\mathbf{y}_1), \mathbf{y}_1) \right) \\
&\quad \cdot \left(h(\mathbf{y}_2) \frac{\partial p_D}{\partial a}(h(\mathbf{y}_2), \mathbf{y}_2) + p_D(h(\mathbf{y}_2), \mathbf{y}_2) \right) \\
&\quad + \left(-1 + \sum_{\mathbf{z} \in Z} \frac{f(\mathbf{z}|\mathbf{y}_1)}{D_{\mathbf{z}}[h]} \right) \left(\frac{\partial}{\partial \delta_{\mathbf{y}_2}} \left(h(\mathbf{y}_1) \frac{\partial p_D}{\partial a}(h(\mathbf{y}_1), \mathbf{y}_1) \right) + \frac{\partial}{\partial \delta_{\mathbf{y}_2}} p_D(h(\mathbf{y}_1), \mathbf{y}_1) \right)
\end{aligned}$$

where we make use of the identity

$$\frac{\delta}{\delta \mathbf{y}} f(h(\mathbf{u}), \mathbf{v}) = \frac{\partial f}{\partial a}(h(\mathbf{u}), \mathbf{v}) \delta_{\mathbf{y}}(\mathbf{u})$$

Equations 40 and 41 directly follow from these results. For, if 40 holds for some $h = h_0$ then

$$\frac{\delta \log L_Z}{\delta \mathbf{y}}[h_0] = 0$$

for all $\mathbf{y}$ and so

$$\log L_{Z_k}[h] - \log L_{Z_k}[h_0] = \frac{1}{2} \frac{\partial^2 \log L_{Z_k}}{\partial (h - h_0)^2}[h_0] + \frac{1}{6} \frac{\partial^3 \log L_{Z_k}}{\partial (h - h_0)^3}[h_0] + \ldots$$

If h is sufficiently near h_0, the second-order term on the right-hand side of the equation dominates all higher-order terms, and the function h_0 corresponds to an absolute maximum if and only if the right-hand side of the equation is always negative. This can be true if and only if equation 41 is true identically.

5. APPENDIX A: MATHEMATICAL PROOFS

5.1. Proof:

We are to show that $\beta_\Gamma(S) = G_\Gamma[\mathbf{1}_S]$. We know that the multi-object density $f_\Gamma(X)$ of the point process Γ is the same thing as the corresponding family of its Janossy densities: $f_\Gamma(\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}) = j_\Gamma(\mathbf{x}_1, \ldots, \mathbf{x}_k)$. So,

$$\begin{aligned}
G_\Gamma[\mathbf{1}_S] &= \sum_{k=0}^{\infty} \frac{1}{k!} \int \mathbf{1}_S(\mathbf{x}_1) \cdots \mathbf{1}_S(\mathbf{x}_k)\, j_\Gamma(\mathbf{x}_1, \ldots, \mathbf{x}_k) d\mathbf{x}_1 \cdots d\mathbf{x}_k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \int_{S^k} f_\Gamma(\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}) d\mathbf{x}_1 \cdots d\mathbf{x}_k = \int_S f_\Gamma(X) \delta X = \beta_\Gamma(S)
\end{aligned}$$

5.2. Proof:

We are to show that

$$\frac{\delta\beta_\Gamma}{\delta\mathbf{y}}(S) = \frac{\delta G_\Gamma}{\delta\mathbf{y}}[\mathbf{1}_S]$$

Let S be a closed set, let $E_\mathbf{x}$ denote an arbitrarily small region surrounding the vector $\mathbf{x}$, and let $\lambda(E_\mathbf{x})$ denote its hypervolume (Lebesgue measure). Our proof will be informal and is based on the observation that, in the limit as $\lambda(E_\mathbf{x}) \to 0$, the characteristic function $\mathbf{1}_{E_\mathbf{x}}$ strongly resembles a Dirac delta function that has been multiplied by a very small postive number:

$$\mathbf{1}_{E_\mathbf{x}}(\mathbf{y}) = \lambda(E_\mathbf{x}) \cdot \frac{\mathbf{1}_{E_\mathbf{x}}(\mathbf{y})}{\lambda(E_\mathbf{x})} \longrightarrow \lambda(E_\mathbf{x}) \cdot \delta_\mathbf{x}(\mathbf{y})$$

Suppose that both the set derivative $\frac{\delta\beta_\Gamma}{\delta\mathbf{y}}$ and the functional derivative $\frac{\delta G_\Gamma}{\delta\mathbf{y}}$ exist. Then by the definition of a set derivative (equation 22, [22, p. 30], [9, pp. 145-146, 150]):

$$\begin{aligned}
\frac{\delta\beta_\Gamma}{\delta\mathbf{y}}(S) &= \lim_{\lambda(E_\mathbf{y})\searrow 0} \frac{\beta_\Gamma(S \cup E_\mathbf{y}) - \beta_\Gamma(S)}{\lambda(E_\mathbf{y})} = \lim_{\lambda(E_\mathbf{y})\searrow 0} \frac{G_\Gamma[\mathbf{1}_{S\cup E_\mathbf{y}}] - G_\Gamma[\mathbf{1}_S]}{\lambda(E_\mathbf{y})} \\
&= \lim_{\lambda(E_\mathbf{y})\searrow 0} \frac{G_\Gamma[\mathbf{1}_S + \mathbf{1}_{E_\mathbf{y}}] - G_\Gamma[\mathbf{1}_S]}{\lambda(E_\mathbf{y})} = \lim_{\lambda(E_\mathbf{y})\searrow 0} \frac{G_\Gamma[\mathbf{1}_S + \lambda(E_\mathbf{y})\delta_\mathbf{y}] - G_\Gamma[\mathbf{1}_S]}{\lambda(E_\mathbf{y})} \\
&= \lim_{\varepsilon\searrow 0} \frac{G_\Gamma[\mathbf{1}_S + \varepsilon\delta_\mathbf{y}] - G_\Gamma[\mathbf{1}_S]}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{G_\Gamma[\mathbf{1}_S + \varepsilon\delta_\mathbf{y}] - G_\Gamma[\mathbf{1}_S]}{\varepsilon} = \frac{\delta G_\Gamma}{\delta\mathbf{y}}[\mathbf{1}_S]
\end{aligned}$$

5.3. Proof:

We are to show that $K(f_{k|k}; \hat{f}_{k|k})$ is minimized if $\hat{D}_{k|k}(\mathbf{x}) = D_{k|k}(\mathbf{x}|Z^{(k)})$. First, notice that

$$\begin{aligned}
K(f_{k|k}; \hat{f}_{k|k}) &= \int f_{k|k}(X|Z^{(k)}) \log f_{k|k}(X|Z^{(k)})\delta X - \int f_{k|k}(X|Z^{(k)}) \log\left(e^{-\hat{N}_{k|k}} \prod_{\mathbf{x}\in X} \hat{D}_{k|k}(\mathbf{x})\right)\delta X \\
&= \text{const.} + \hat{N}_{k|k} - \int \left(f_{k|k}(X|Z^{(k)}) \sum_{\mathbf{x}\in X} \log \hat{D}_{k|k}(\mathbf{x})\right)\delta X
\end{aligned}$$

However, a bit of algebra shows that

$$\int \left(f_{k|k}(X|Z^{(k)}) \sum_{\mathbf{x}\in X} \log \hat{D}_{k|k}(\mathbf{x})\right)\delta X = N_{k|k} \int s_{k|k}(\mathbf{x}|Z^{(k)}) \log \hat{s}_{k|k}(\mathbf{x}) d\mathbf{x} + N_{k|k} \log \hat{N}_{k|k}$$

where $s_{k|k}(\mathbf{x}|Z^{(k)}) = D_{k|k}(\mathbf{x}|Z^{(k)})/N_{k|k}$ and $\hat{s}_{k|k}(\mathbf{x}|Z^{(k)}) = \hat{D}_{k|k}(\mathbf{x}|Z^{(k)})/\hat{N}_{k|k}$. So,

$$\begin{aligned}
K(f_{k|k}; \hat{s}_{k|k}) &= \text{const.} - N_{k|k} \int s_{k|k}(\mathbf{x}|Z^{(k)}) \log \hat{s}_{k|k}(\mathbf{x}) d\mathbf{x} + \hat{N}_{k|k} - N_{k|k} \log \hat{N}_{k|k} \\
&= \text{const.} + N_{k|k} K(s_{k|k}, \hat{s}_{k|k}) + \hat{N}_{k|k} - N_{k|k} \log \hat{N}_{k|k}
\end{aligned}$$

This is minimized when $K(s_{k|k}, \hat{s}_{k|k})$ and $\hat{N}_{k|k} - N_{k|k} \log \hat{N}_{k|k}$ are separately minimized, i.e. when $s_{k|k} = \hat{s}_{k|k}$ and $\hat{N}_{k|k} = N_{k|k}$.

5.4. Proof

We are given a sensor with single-target measurement model $\mathbf{z} = \mathbf{g}(\mathbf{x}) + \mathbf{V}$, state space $\mathbb{S}$, likelihood function $f(\mathbf{z}|\mathbf{x}) = f_{\mathbf{V}}(\mathbf{z} - \mathbf{g}(\mathbf{x}))$, state-dependent probability of detection $p_D(\mathbf{x})$, and a Poisson track-cluster Γ with intensity function $D(\mathbf{y})$ where $a = \int D(\mathbf{y})d\mathbf{y} > 0$ is the expected number of targets. Suppose further that Γ is a Poisson point track-cluster and that the probability of detection $p_D(a, \mathbf{x})$ is a function of average target number a as well as of target state. We are to show that if the track-cluster is a point track-cluster, then the observation-process generated by it is a Poisson process with intensity function $D(\mathbf{z}|a, \mathbf{x}) = a \cdot p_D(a, \mathbf{x}) \cdot f(\mathbf{z}|\mathbf{x})$.

Let ν be a Poisson-distributed random non-negative integer with parameter a and let $\mathbf{X}_1, ..., \mathbf{X}_i, ...$ be independent, identically distributed (i.i.d.) random vectors with distribution $s(\mathbf{x}) = D(\mathbf{x})/a$. From the footnote in section 3.2, we know that $\Gamma = \{\mathbf{X}_1, ..., \mathbf{X}_\nu\}$ is a Poisson track-cluster with the specified characteristics. Let $\emptyset_i^q$ be a random set which takes only the two values $\emptyset, \mathbb{S}$ with $\Pr(\emptyset_i^q = \emptyset) = 1 - q$ and $\Pr(\emptyset_i^q = \mathbb{S}) = q$. Then the observation-set Σ produced by the track-cluster is

$$\Sigma = \bigcup_{i=1}^{v} \left(\{\mathbf{g}(\mathbf{X}_i) + \mathbf{V}_i\} \cap \emptyset_i^{p_D(a, \mathbf{X}_i)} \right)$$

where $\mathbf{V}_1, ..., \mathbf{V}_i, ...$ are i.i.d. The corresponding belief-mass function is

$$\begin{aligned}
\beta_\Sigma(S) &= \Pr(\Sigma \subseteq S) = \sum_{n=0}^{\infty} p_\nu(n) \prod_{i=1}^{n} \Pr\left(\{\mathbf{g}(\mathbf{X}_i) + \mathbf{V}_i\} \cap \emptyset_i^{p_D(a, \mathbf{X}_i)} \subseteq S \right) \\
&= \sum_{n=0}^{\infty} p_\nu(n) \prod_{i=1}^{n} \Pr\left(\{\mathbf{g}(\mathbf{X}_i) + \mathbf{V}_i\} \cap \emptyset_i^{p_D(a, \mathbf{X}_i)} \subseteq S \right)
\end{aligned}$$

However, assuming independence of all random quantities we get

$$\begin{aligned}
\Pr\left(\{\mathbf{g}(\mathbf{X}_i) + \mathbf{V}_i\} \cap \emptyset_i^{p_D(a, \mathbf{X}_i)} \subseteq S \right) &= \int \Pr\left(\{\mathbf{g}(\mathbf{y}) + \mathbf{V}_i\} \cap \emptyset_i^{p_D(a, \mathbf{y})} \subseteq S \right) s(\mathbf{x}) d\mathbf{x} \\
&= \int (1 - p_D(a, \mathbf{y}) + p_D(a, \mathbf{y}) \Pr(\mathbf{g}(\mathbf{y}) + \mathbf{V}_i \in S))\, s(\mathbf{y}) d\mathbf{y} \\
&= \int (1 - p_D(a, \mathbf{y}) + p_D(a, \mathbf{y})\, p_{\mathbf{Z}}(S|\mathbf{y}))\, s(\mathbf{y}) d\mathbf{y}
\end{aligned}$$

where $p_{\mathbf{Z}}(S|\mathbf{y}) = \int_S f(\mathbf{z}|\mathbf{y}) d\mathbf{z}$. So,

$$\begin{aligned}
\beta_\Sigma(S) &= \sum_{n=0}^{\infty} p_\nu(n) \left(\int (1 - p_D(\mathbf{y}) + p_D(\mathbf{y})\, p_{\mathbf{Z}}(S|\mathbf{y}))\, s(\mathbf{y}) d\mathbf{y} \right)^n \\
&= \exp\left(a \int (1 - p_D(\mathbf{y}) + p_D(\mathbf{y})\, p_{\mathbf{Z}}(S|\mathbf{y}))\, s(\mathbf{y}) d\mathbf{y} - a \right)
\end{aligned}$$

This is the belief-mass function of a Poisson observation-cluster. Now assume that the track-cluster is a point cluster, i.e. $s(\mathbf{y}) = \delta_{\mathbf{x}}(\mathbf{y})$. Then the belief-mass function becomes

$$\beta_\Sigma(S|a, \mathbf{x}) = \exp(a p_D(a, \mathbf{x}) (p_{\mathbf{Z}}(S|\mathbf{x}) - 1))$$

This is the belief-mass function of a Poisson process with intensity function $D(\mathbf{z}|a, \mathbf{x}) = a \cdot p_D(a, \mathbf{x}) \cdot f(\mathbf{z}|\mathbf{x})$, as claimed.

5.5. Proof:

We are to show that

$$f(Z|\mathcal{X}) = e^{-a_{\mathcal{X}}} \prod_{\mathbf{z}\in Z} D(\mathbf{z}|\mathcal{X}), \qquad D(\mathbf{z}|\mathcal{X}) = \sum_{i=1}^{n} a_i p_D(a_i, \mathbf{x}_i) f(\mathbf{z}|\mathbf{x}_i), \qquad a_{\mathcal{X}} = \sum_{i=1}^{n} a_i p_D(a_i, \mathbf{x}_i)$$

The belief-mass function of each point-Poisson process is

$$\beta_i(S) = \exp\left(a_i p_D(a_i, \mathbf{x}_i) p_{\mathbf{Z}}(S|\mathbf{x}_i) - a_i p_D(a_i, \mathbf{x}_i)\right)$$

Since the processes are independent, the belief-mass function of their union is also that of a Poisson process:

$$\beta(S) = \beta_1(S) \cdots \beta_n(S) = \exp\left(\sum_{i=1}^{n} a_i p_D(a_i, \mathbf{x}_i) p_{\mathbf{Z}}(S|\mathbf{x}_i) - \sum_{i=1}^{n} a_i p_D(a_i, \mathbf{x}_i)\right) = \exp\left(D(\mathbf{z}|\mathcal{X}) - a_{\mathcal{X}}\right)$$

as claimed.

5.6. Proof

We are to show that

$$\begin{aligned}
\frac{\delta D_{\mathbf{z}}}{\delta \mathbf{y}}[h] &= \left(h(\mathbf{y}) \frac{\partial p_D}{\partial a}(h(\mathbf{y}), \mathbf{y}) + p_D(h(\mathbf{y}), \mathbf{y})\right) f(\mathbf{z}|\mathbf{y}) \\
\frac{\delta \alpha}{\delta \mathbf{y}}[h] &= h(\mathbf{y}) \frac{\partial p_D}{\partial a}(h(\mathbf{y}), \mathbf{y}) + p_D(h(\mathbf{y}), \mathbf{y}) \\
\frac{\delta \log L_Z}{\delta \mathbf{y}}[h] &= \left(-1 + \sum_{\mathbf{z}\in Z} \frac{f(\mathbf{z}|\mathbf{y})}{D_{\mathbf{z}}[h]}\right) \left(h(\mathbf{y}) \frac{\partial p_D}{\partial a}(h(\mathbf{y}), \mathbf{y}) + p_D(h(\mathbf{y}), \mathbf{y})\right)
\end{aligned}$$

The second equation follows easily from the first, and the third follows from the first and second since

$$\frac{\delta \log L_Z}{\delta \mathbf{y}}[h] = -\frac{\delta}{\delta \mathbf{y}} \alpha[h] + \sum_{\mathbf{z}\in Z} \frac{\delta}{\delta \mathbf{y}} \log D_{\mathbf{z}}[h]$$

. So, we need only prove the first equation:

$$\begin{aligned}
\frac{\delta D_{\mathbf{z}}}{\delta \mathbf{y}}[h] &= \lim_{\varepsilon\to 0} \frac{D_{\mathbf{z}}[h + \varepsilon \delta_{\mathbf{y}}] - D_{\mathbf{z}}[h]}{\varepsilon} = \int \left(\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \begin{array}{c} (h + \varepsilon\delta_{\mathbf{y}})(\mathbf{x}) p_D(h(\mathbf{x}) + \varepsilon\delta_{\mathbf{y}}(\mathbf{x}), \mathbf{x}) f(\mathbf{z}|\mathbf{x}) \\ -h(\mathbf{x}) p_D(h(\mathbf{x}), \mathbf{x}) f(\mathbf{z}|\mathbf{x}) \end{array}\right) d\mathbf{x} \\
&= \int \left(\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \begin{array}{c} h(\mathbf{x}) p_D(h(\mathbf{x}) + \varepsilon\delta_{\mathbf{y}}(\mathbf{x}), \mathbf{x}) f(\mathbf{z}|\mathbf{x}) + \varepsilon\delta_{\mathbf{y}}(\mathbf{x}) p_D(h(\mathbf{x}) + \varepsilon\delta_{\mathbf{y}}(\mathbf{x}), \mathbf{x}) f(\mathbf{z}|\mathbf{x}) \\ -h(\mathbf{x}) p_D(h(\mathbf{x}), \mathbf{x}) f(\mathbf{z}|\mathbf{x}) \end{array}\right) d\mathbf{x} \\
&= \int \left(\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \begin{array}{c} h(\mathbf{x}) p_D(h(\mathbf{x}) + \varepsilon\delta_{\mathbf{y}}(\mathbf{x}), \mathbf{x}) f(\mathbf{z}|\mathbf{x}) - h(\mathbf{x}) p_D(h(\mathbf{x}), \mathbf{x}) f(\mathbf{z}|\mathbf{x}) \\ +\varepsilon\delta_{\mathbf{y}}(\mathbf{x}) p_D(h(\mathbf{x}) + \varepsilon\delta_{\mathbf{y}}(\mathbf{x}), \mathbf{x}) f(\mathbf{z}|\mathbf{x}) \end{array}\right) d\mathbf{x}
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{\delta D_{\mathbf{z}}}{\delta \mathbf{y}}[h] &= \int h(\mathbf{x}) f(\mathbf{z}|\mathbf{x}) \left(\lim_{\varepsilon\to 0} \frac{p_D(h(\mathbf{x}) + \varepsilon\delta_{\mathbf{y}}(\mathbf{x}), \mathbf{x}) - p_D(h(\mathbf{x}), \mathbf{x})}{\varepsilon}\right) d\mathbf{x} \\
&\quad + \int \lim_{\varepsilon\to 0} \delta_{\mathbf{y}}(\mathbf{x})\; p_D(h(\mathbf{x}) + \varepsilon\delta_{\mathbf{y}}(\mathbf{x}), \mathbf{x})\; f(\mathbf{z}|\mathbf{x}) d\mathbf{x} \\
&= \int h(\mathbf{x}) f(\mathbf{z}|\mathbf{x}) \delta_{\mathbf{y}}(\mathbf{x}) \frac{\partial p_D}{\partial a}(h(\mathbf{x}), \mathbf{x}) d\mathbf{x} + \int \delta_{\mathbf{y}}(\mathbf{x})\; p_D(h(\mathbf{x}), \mathbf{x})\; f(\mathbf{z}|\mathbf{x}) d\mathbf{x} \\
&= \left(h(\mathbf{y}) \frac{\partial p_D}{\partial a}(h(\mathbf{y}), \mathbf{y}) + p_D(h(\mathbf{y}), \mathbf{y})\right) f(\mathbf{z}|\mathbf{y})
\end{aligned}$$

6. CONCLUSIONS

In this paper I introduced multitarget moment statistics of arbitrary order, and have tried to illustrate their potential importance for multitarget tracking. I emphasized the importance of point process theory to multisensor-multitarget detection, tracking, and identification. Despite its neglect in multitarget tracking circles, this theory is the statistical foundation for all multi-object problems, and its importance to future theoretical and practical research should not be underestimated. The point process literature contains a vast array of results, techniques, and computational approaches that await "mining" for application to multitarget problems. Towards that end, I have provided a brief but systematic introduction to point process theory, and in particular to the "engineering-friendly" version of it that I call finite-set statistics (FISST).

I showed how a viable Bayes nonlinear filter can be devised for the first-order multitarget moment (the "probability hypothesis density" or "PHD"), thus yielding a computable, systematic multitarget tracking approach that avoids report-to-track association. I described preliminary simulations that suggest its possible benefits and possible challenges in practice. Chief among the limitations, other than the necessity for high signal-to-noise ratio, is the need for good multi-peak extraction algorithms. This is because multi-peak extraction largely replaces report-to-track association in a PHD multitarget filter. I also pointed out an open research question: the problem of defining a *track-valued* (rather than density- or measure-valued) first-order multitarget moment statistics.

I initiated a study of the second-order statistics of multitarget systems. While this study is preliminary, it suggests possible directions for research. I described the second-order moment and covariance statistics of a multitarget system, and indicated why it does not appear possible to develop a computationally tractable multitarget tracking filter based on these statistics at this time. However, I also described a potential path to a multitarget generalization of the Kalman filter. This approach depends upon (1) being able to extend any multitarget likelihood function to a so-called "multitarget log-likelihood functional," and (2) expanding this functional in a Taylor's series, using a generalization of the FISST set derivative called the functional derivative.

ACKNOWLEDGMENTS

The work on the implementation of the PHD, described in section 3.4, was supported by Lockheed Martin IR&D and by the U.S. Air Force Research Laboratory (AFRL) under contract F30602-00-C-0107. The work on the Hausdorff "multitarget miss distance" also reported in that section was supported by AFRL under contract F30602-98-C-0270. Special thanks to Dr. Tim Zajic and Dr. John Hoffman of Lockheed Martin Tactical Systems and Dr. Adel El-Fellah of Scientific Systems Co., Inc. for their contributions to this work. All other work described in this paper was supported by the U.S. Army Research Office under contract DAAG55-98-C-0039. The content does not necessarily reflect the position or policy of the Government. No official endorsement should be inferred.

REFERENCES

1. M. Bardin, "Multidimensional Point Processes and Random Closed Sets," *J. Applied Prob.* **21**, pp. 173-178, 1984
2. Y. Bar-Shalom and X.-R. Li, *Estimation and Tracking: Principles, Techniques, and Software*, Artech House, 1993
3. R.E. Bethel and G.J. Paras, "A PDF multitarget-tracker," *IEEE Trans. Aerospace and Electronics Sys.* **30:2**, pp. 386-403, 1994
4. G. Casella and R.L. Berger, *Statistical Inference*, Wadsworth & Brooks, 1990
5. C.K. Chui and G. Chen, *Kalman Filtering with Real-Time Applications,* Third Edition, Springer, 1999
6. D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Springer-Verlag, 1988
7. A. El-Fallah, T. Zajic, R. Mahler, B. Lazja-Rooks, and R. Mehra, "Multitarget nonlinear filter based on spectral compression and probability hypothesis densities," in *Signal Processing, Sensor Fusion, and Target Recognition X*, I. Kadar, ed., *SPIE Proc.* **4380**, 2001, to appear
8. I.M. Gel'fand and A.M. Yaglom, "Integration in Function Spaces and its Applications in Quantum Physics," *J. of Math. Physics* **1:1**, pp. 48-560, 1960

9. I.R. Goodman, R.P.S. Mahler, and H.T. Nguyen, *Mathematics of Data Fusion*, Kluwer Academic Publishers, 1997
10. J. Goutsias, R. Mahler, and H.T. Nguyen, eds., *Random Sets: Theory and Application*, Springer-Verlag, 1997
11. R.A. Iltis, "State Estimation Using an Approximate Reduced Statistics Algorithm," *IEEE Trans. Aerospace and Electr. Sys.* **35:4**, pp. 1161-1172, 1999
12. R. Kulhavý, "Recursive Nonlinear Estimation: A Geometric Approach," *Automatica* **26:3**, pp. 545-555, 1990
13. Y.C. Ho and R.C.K. Lee, "A Bayesian Approach to Problems in Stochastic Estimation and Control," *IEEE Trans. Automatic Contr.*, **9**, pp. 333-339, 1964
14. A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970
15. A.F. Karr, *Point Processes and Their Statistical Inference*, Second Edition, Marcel Dekker, 1991
16. M. Kouritzin, "Particle Approximations," presentation at the AFRL/AFOSR Workshop on Nonlinear Filtering Methods for Tracking, Dayton OH, Feb. 21-22 2001
17. R. Mahler, "Approximate Multisensor-Multitarget Joint Detection, Tracking, and Identification Using a First-Order Multitarget Moment Statistic," submitted for publication
18. R. Mahler, "Global Integrated Data Fusion," *Proc. 7'th Nat'l Symp. on Sensor Fusion*, Vol. I (Unclassified), Mar. 16-18 1994, Sandia National Laboratories, Albuquerque, ERIM, Ann Arbor MI, pp. 187-199
19. R. Mahler, "Information for fusion management and performance estimation," in *Signal Processing, Sensor Fusion, and Target Recognition VII*, I. Kadar, ed., *SPIE Proc.* **3374**, pp. 64-75, 1998
20. R. Mahler, "Information Theory and Data Fusion," *Proc. 8'th Nat'l Symp. on Sensor Fusion*, Vol. I (Unclassified), Texas Instruments, Dallas, Mar. 17-19 1995, ERIM, Ann Arbor MI, pp. 279-292
21. R. Mahler, "Engineering Statistics for Multitarget Tracking," IEEE Workshop on Multi-Object Tracking, July 8 2001, *Proc. 2001 IEEE Int'l Conf. on Computer Vision*, Vancouver
22. R. Mahler, *An Introduction to Multisource-Multitarget Statistics and Its Applications*, Lockheed Martin Technical Monograph, 104 pages, March 15 2000
23. R. Mahler, "Multisensor-Multitarget Statistics," in *Data Fusion Handbook*, D. Hall, ed., CRC Press, Boca Raton, 2001, to appear
24. R. Mahler, "Multisource-multitarget filtering: A unified approach," in *Signal and Data Processing of Small Targets 1998*, O.E. Drummond, ed., *SPIE Proc.* **3373**, pp. 296-307, 1998
25. R. Mahler, "Multitarget markov motion models," in *Signal Processing, Sensor Fusion, and Target Recognition VIII*, I. Kadar, ed., *SPIE Proc.* **3720**, pp. 47-57, 1999
26. R. Mahler, "The Search for Tractable Bayes Multitarget Filters," in *Signal and Data Processing of Small Targets 2000*, O.E. Drummond, ed., *SPIE Proc.* **4048**, pp.310-320, 2000
27. R. Mahler, "A Theoretical Foundation for the Stein-Winter 'Probability Hypothesis Density (PHD)' Multitarget Tracking Approach," *Proc. 2000 MSS Nat'l Symp. on Sensor and Data Fusion*, Vol. I (Unclassified), San Antonio TX, June 20-22 2000, pp. 99-118
28. R. Mahler, "What is the 'Multitarget Average Value' of a Multitarget Scenario," presentation at GTRI/ONR Workshop on Tracking and Filtering, Georgia Tech Research Institute, Marietta GA, June 1 1998
29. R. Mahler and T. Zajic,"Multitarget filtering using a multitarget first-order moment statistic," in *Signal Processing, Sensor Fusion, and Target Recognition X*, I. Kadar, ed., *SPIE Proc.* **4380**, 2001, to appear
30. G. Mathéron, *Random Sets and Integral Geometry*, Wiley, 1975
31. E.W. Montroll, "Markoff Chains, Wiener Integrals, and Quantum Theory," *Comm. on Pure & Applied Math.* **5**, pp. 415-453, 1952
32. S. Mori, "Random Sets in Data Fusion: Multi-Object State-Estimation as a Foundation of Data Fusion Theory," in *Random Sets: Theory and Applications*, J. Goutsias, R.P.S. Mahler, and H.T. Nguyen, eds., Springer-Verlag, pp. 185-207, 1997
33. S. Mori, "A Theory of Informational Exchanges–Random Set Formalism," *Proc. 1998 IRIS Nat'l Symp. on Sensor and Data Fusion*, Vol. I (Unclassified), Marietta GA, March 31-March 2 1998, ERIM, pp. 147-158
34. S. Mori, "Random Sets in Data Fusion: Formalism to New Algorithms," *Proc. Third Int'l Conf. on Information Fusion*, Paris, July 10-13 2000, to appear
35. S. Mori, C.-Y. Chong, E. Tse, and R.P. Wishner, "Tracking and Classifying Multiple Targets Without A Priori Identification," *IEEE Trans. Auto. Contr.* **31:5**, pp. 401-409, 1986
36. J.E. Moyal, "The general theory of stochastic population processes," *Acta Mathematica* **108**, pp. 1-31, 1962

37. S. Musick, K. Kastella, and R. Mahler, "A practical implementation of joint multitarget probabilities," in *Signal Processing, Sensor Fusion, and Target Recognition VII*, I. Kadar, ed., *SPIE Proc.* **3374**, pp. 26-37, 1998
38. N. Portenko, H. Salehi, and A. Skorokhod, "On optimal filtering of multitarget tracking systems based on point processes observations," *Random Operators and Stochastic Equations* **1**, pp. 1-34, 1997
39. B.D. Ripley, "Locally finite random sets: foundations for point process theory," *Annals of Prob.* **4:6**, pp. 983-994, 1976
40. L.H. Ryder, *Quantum Field Theory*, 2nd Edition, Cambridge U. Press, 1996
41. A. Skorokhod, "Optimal filtering in target tracking in the presence of uniformly distributed errors and false targets," June 1 1998, 1998 GTRI/ONR Workshop on Target Tracking and Sensor Fusion, Georgia Tech Research Institute, Marietta GA
42. H.W. Sorenson, "Recursive Estimation for Nonlinear Dynamic Systems," in *Bayesian Analysis of Statistical Time Series and Dynamic Models*, J.C. Spall, ed., Marcel Dekker, 1988
43. D.L. Snyder and M.I. Miller, *Random Point Processes in Time and Space*, Second Edition, Springer, 1991
44. A. Srivastava, M.I. Miller, and U. Grenander, "Jump-Diffusion Processes for Object Tracking and Direction Finding," *Proc. 29th Allerton Conf. on Communication, Control, and Computing*, U. of Illinois, Urbana, pp. 563-570, 1991
45. A. Srivastava, N. Cutaia, M.I. Miller, J.A. O'Sullivan, and D.L. Snyder, "Multitarget Narrowband Direction Finding and Tracking Using Motion Dynamics," *Proc. 30th Allerton Conf. on Communication, Control, and Computing*, Sept. 30-Oct.2 1992, Monticello IL, pp. 279-288, 1992
46. M.C. Stein and C.L. Winter, "An Additive Theory of Probabilistic Evidence Accrual," Los Alamos National Laboratories Report LA-UR-93-3336, 1993
47. L.D. Stone, C.A. Barlow, and T.L. Corwin, *Bayesian Multiple Target Tracking*, Artech House, 1999
48. D. Stoyan, W.S. Kendall, and J. Meche, *Stochastic Geometry and Its Applications*, Second Edition, John Wiley & Sons, 1995
49. K.M. Tao, R. Abileah, and J.D. Lawrence, "Multiple-target tracking in dense noisy environments: a probabilistic mapping perspective," in *Signal and Data Processing of Small Targets 2000*, O.E. Drummond, ed., *SPIE Proc.* **4048**, pp. 474-485, 2000
50. H.L. van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*, John Wiley & Sons, 1968
51. R.B. Washburn, "A Random Point Process Approach to Multi-Object Tracking," *Proc. Amer. Contr. Conf.* **3**, Minneapolis, pp. 1846-1852, June 10-12 1987
52. T. Zajic, J. Hoffman, and R. Mahler, "Scientific performance metrics for data fusion: new results," in *Signal Processing, Sensor Fusion, and Target Recognition IX*, I. Kadar, ed., *SPIE Proc.* **4052**, pp. 172-182, 2000